الحياة 3.0

الكون بشراً في عصر الذكاء الاصطناعي

ماكس تيغمارك



ترجمة: د. ليلى الموسوي

الأكثر مبيعاً في قائمة نيويورك تايمز

الحياة 3.0

كيف سيؤثر الذكاء الاصطناعي في الجريمة، العدالة، الوظائف، وحتى معني كوننا بشرا؟ إن صعود الذكاء الاصطناعي قادر على أن يغيّر مستقبلنا أكثر من أي تقنية أخرى- ولا أحد مهيأ أفضل من تيغ ماكس، الأستاذ بمعهد ماساتشوستيس للتكنولوجيا MIT الذي ساعد التيار البحثي العام على التوصل إلى كيفية إبقاء الذكاء الاصطناعي مفيدا.

أنىّ لنا أن ندفع نحو تنامي ازدهارنا عبر الأتمتة من دون أن نخلف البشر في عوز مالي أو من دون هدف؟ ما النصيحة الوظيفية التي نسديها لأطفال اليوم؟ وكيف نبني أنظمة ذكاء اصطناعي صلدة، بحيث تؤدي ما نريد لها أن تؤديه من دون أن تنهار، أن تصاب بعطب، أو تتعرض للقرصنة؟ هل يتعين علينا أن نخشى اندلاع سباق تسلح لتطوير الأسلحة القاتلة ذاتية التحكم؟ وفي النهاية هل ستصبح الآلات أفضل منا في تنفيذ جميع المهام، وتحل محل البشر في سوق العمل أو ربما تحل محلنا بالمطلق؟ هل سيساعد الذكاء الاصطناعي على ازدهار الحياة كما لم يسبق لها من قبل أو يمنحنا قوة أكبر من طاقتنا؟

ما نوع المستقبل الذي تريده؟ هذا الكتاب يمكنك من أن تشارك في المحادثة الأهم في وقتنا هذا. فهو يعرض الرؤى من كامل مدى وجهات النظر المتباينة، ولا يحيد عن مناقشة القضايا الجدلية- من الذكاء الفائق إلى المعني، ومن الوعي إلى الحدود المادية القصوى التي تحد الحياة والكوزموس.

ISBN
9 789921 719307

KFAS | إحدى شركات Company | التقدم العلمي للنشر Advancement of Science Publishing

حياة
3.0

الكون بشراً في عصر الذكاء الاصطناعي

ماكس تيغمارك

ترجمة: د. ليلى الموسوي

الطبعة الأولى - الكويت 2019

فريق عمل النسخة العربية

رئيس مجلس الإدارة

د. سلام أحمد العبلاني

نائب رئيس مجلس الإدارة والرئيس التنفيذي

د. ليلى الموسوي

الترجمة العربية

د. ليلى الموسوي

هيئة التحرير

د.عبدالله بدران

عبدالله المهنا

محمد الحسن

مي بورسلي

ريهام العوضي

التسويق

خالد الرشيدي

التدقيق اللغوي

فادي بدارنه

الغرافيك والتنضيد

خالد كلارجي

سكينة عبد الصمد

المتابعة والتنسيق

دانيا حداد

هذا الكتاب المترجمَ يعبر عن وجهة نظر المؤلف ودار النشر،
ولا تتحمل مؤسسة الكويت للتقدم العلمي أي مسؤولية أو تبعات عن مضمون الكتاب.

مؤسسة الكويت للتقدم العلمي

التقدم العلمي للنشر Advancement of Science Publishing | احدى شركات Company | KFAS

الطبعة الأولى - الكويت 2019

ISBN: 978-9921-719-30-7

بسم الله الرحمن الرحيم

إهداء

إلى فريق معهد مستقبل الحياة
الذي جعل كل شيء ممكناً

Life 3.0
By
Max Tegmark

المحتويات

شكر وتقدير

أنا ممتن حقّاً لكل من شجعني وساعدني على كتابة هذا الكتاب، بمن في ذلك

عائلتي وأصدقائي ومدرسيّ وزملائي وشركائي في الأبحاث لدعمهم وإلهامهم لي على مرّ السنين.

أمي لتذكية فضولي حول الوعي والمعنى، أبي لروحه القتالية لجعل العالم مكاناً أفضل.

ولداي، فيليب وألكساندر، لعرضهما عجائب ذكاء المستوى البشري الناشئ.

جميع المتحمسين للعلوم والتكنولوجيا في جميع أنحاء العالم الذين اتصلوا بي على مرّ السنين للسؤال والتعليق والتشجيع على تعقب أفكاري ونشرها، وكيل أعمالي، جون بروكمان، لليّه ذراعي حتى وافقت على كتابة هذا الكتاب.

بوب بينا وجيسي ثالر وجيريمي إنغلاند للمناقشات المفيدة حول النجوم الزائفة والسفاليرونSphaleron والديناميكا الحرارية على التوالي.

أولئك الذين قدموا لي تعليقات على أجزاء من المخطوطة، بمن في ذلك أمي، أخي بير، لويزا بايت، روب بينزينغر، كاترينا بيرغستروم، إريك برينغولفسون، دانييلا تشيتا، ديفيد تشالمرز، نعمة ديغاني، هنري لين، إلين مالمسكولد، توبي أورد، جيرمي أوين، لوكاس بيري، أنثوني روميريو، نيت سوارس، ويان تالين.

الأبطال الخارقون الذين علّقوا على مسودات الكتاب بكامله وهم: ميا، والدي، أنثوني أوغيير، بول آلموند، ماثيو غريفس، فيليب هيلبيغ، ديفيد ماربل، هاوارد ميسينغ، لينيو سيوان، ماريا سولياسيس، ومحرري دان فرامك، وعلى الأخص

ميا، مُلهمتي الحبيبة ورفيقة سفري، لتشجيعها الأزلي ودعمها وإلهامها، والتي من دونها ما كان هذا الكتاب ليجد طريقه نحو الصدور.

حياة 3.0

هل تعتقد أن ذكاءً اصطناعياً يفوق البشر سيُطور في هذا القرن؟

نعم

اقلب إلى الصفحة التالية

لا

اقلب إلى الفصل الأول (صفحة 31)

مقدمة

—

حكاية فريق أوميغا

كان فريق أوميغا هو روح الشركة. فبينما جلبت بقية الشركة الأموال اللازمة للحفاظ على سير الأعمال، فإن فريق أوميغا -من خلال تطبيقات تجارية متنوعة للذكاء الاصطناعي المتخصص- ظلّ يدفع سعيهم وراء ما كان دائماً حلم أي رئيس تنفيذي: بناء نظام ذكاء اصطناعي عام. ومعظم الموظفين الآخرين نظروا إلى «الأوميغا»، -كما يطلقون عليهم تَودُّداً- كمجموعة من الحالمين الغارقين في الأفكار الخيالية، وفعليا على بُعد عقود من تحقيق هدفهم. ولكنهم وفروا لهم -بسعادة- ما يريدون؛ لأنهم معجبون بالمكانة التي تسبغها أعمال الفريق أوميغا المتطورة على شركتهم، كما أنهم يثمّنون الخوارزميات المحسنة التي قدمها الأوميغا.

غير أن ما لم يدركوه هو أن فريق أوميغا صاغ صورته هذه بعناية لإخفاء سره: فقد كانوا قريبين جدّاً من تنفيذ أكثر الخطط جرأة في تاريخ البشرية. وقد اختارهم الرئيس التنفيذي ذو الشخصية الكاريزمية ليس فقط لكونهم باحثين بارعين، ولكن أيضاً لطموحهم ورؤيتهم المثالية والالتزام الشديد بمساعدة البشر. وذكّرهم بأن خطتهم كانت خطيرة جدّاً، وأنه إذا اكتشفت الحكومات القوية ذلك، فإنها ستفعل أي شيء تقريباً - بما في ذلك الاختطاف - لإيقاف الخطة، أو لتنفيذ الخيار الذي يفضلونه: سرقة الكود Code. غير أن أفراد الفريق كانوا جميعاً ملتزمين مئة في المئة، للسبب نفسه الذي جعل العديد من كبار علماء الفيزياء في العالم ينضمون إلى مشروع مانهاتن لتطوير الأسلحة النووية: كانوا مقتنعين بأنه إذا لم يفعلوا ذلك أولاً، فإن شخصاً أقل مثالية سيفعل ذلك.

وظل الذكاء الاصطناعي الذي طوّروه، وأطلقوا عليه تحببا اسم بروميثيوس Prometheus، يزداد مقدرة. وعلى الرغم من أن قدراته المعرفية لا تزال أدنى كثيراً من قدرات البشر في العديد من المجالات، مثلاً: المهارات الاجتماعية، فإن فريق أوميغا جهد لجعله غير عادي في مهمة واحدة بعينها: برمجة نظم الذكاء الاصطناعي، فقد اختاروا هذه الاستراتيجية عن قصد؛ لأنهم اقتنعوا بحجة تفجّر الذكاء التي قال بها عالم الرياضيات البريطاني إيرفينغ غود Irving Good في عام 1965: «دعونا نُعرّف آلة فائقة الذكاء كآلة

يمكنها أن تفوق كل الأنشطة الفكرية لأي فرد مهما كان ذكيّاً، ولما كان تصميم الآلات هو أحد هذه الأنشطة الفكرية، فإن آلة فائقة الدقة يمكنها تصميم آلات أفضل، عندئذ سيكون هناك بلا شك 'ذكاء متفجر' Intelligence explosion، وسيترك وراءنا ذكاء الإنسان. ومن ثم، فإن أول آلة فائقة الذكاء هي آخر اختراع يحتاج إليه الإنسان، شريطة أن يكون الجهاز مُهادِناً بما فيه الكفاية ليخبرنا بكيفية إبقائه تحت السيطرة».

وأدرك فريق أوميغا أنه إذا تمكنوا من الحصول على هذا التحسين الذاتي المستمر؛ فسرعان ما يصبح الجهاز ذكيّاً بما فيه الكفاية حتى يتمكن من تعليم نفسه جميع المهارات البشرية الأخرى التي ستكون مفيدة له.

الملايين الأولى

كانت الساعة التاسعة صباح الجمعة عندما قرروا إطلاقه، كان بروميثيوس «يُحمِحم» منهمكاً بالعمل عبر مجموعة أجهزته الحاسوبية المصممة خصيصاً، والتي كانت تقبع على صفوف طويلة من الرفوف في غرفة واسعة مكيفة الهواء يمكن التحكم في الدخول إليها. ولأسباب أمنية، فقد كانت مفصولة كليّةً عن الإنترنت، ولكنها تحتوي على نسخة محلية من جزء كبير من الويب (ويكيبيديا، ومكتبة الكونغرس، وتويتر، ومجموعة مختارة من موقع يوتيوب، والكثير من فيسبوك، وما إلى ذلك) لاستخدامها كبيانات تدريبية للتعلم منها*. وقد اختاروا وقت البدء هذا للعمل من دون أي إزعاج: فقد اعتقدت عائلاتهم وأصدقاؤهم أنهم كانوا في مُعتَكف للشركة في عطلة نهاية الأسبوع، وتَمَّ إمداد المطبخ الصغير بالمأكولات التي تحضر بالميكروويف ومشروبات الطاقة، وكانوا على استعداد لإطلاقه.

عندما أطلقوه، كان بروميثيوس أسوأ منهم بقليل من حيث قدرته على برمجة نظم الذكاء الاصطناعي، لكنه عوض عن ذلك من خلال كونه أسرع بكثير، فأنفق ما يعادل الآلاف من سنوات الأفراد في حل المشكلة في وقت لم يتعدَّ الوقت الذي اجترع فيه الفريق مشروب الطاقة ريد بول. وبحلول الساعة العاشرة صباحاً، استكملت عملية إعادة التصميم الأولى لنفسها، الإصدار 2.0، التي كانت أفضل قليلاً لكنها ما زالت دون المستوى البشري. وبحلول الوقت الذي أطلق فيه بروميثيوس الإصدار 5.0 في الساعة 2 بعد الظهر، كان الفريق أوميغا مصعوقاً من الدهشة: فقد دفع بروميثيوس بمعايير الأداء إلى مستويات عالية جدّاً، وبدا أن معدل التقدم يتسارع، وبحلول الليل، قرروا تطبيق بروميثيوس 10.0 لبدء المرحلة الثانية من خطتهم: كسب المال.

* للتبسيط، استخدمت حيثيات الاقتصاد والتكنولوجيا في يومنا هذا لكتابة هذه القصة، على الرغم من أن معظم الباحثين يعتقدون أن الذكاء الاصطناعي العام على المستوى البشري لا يزال بعيداً عنَّا لعقود. فيجب أن تصبح خطة أوميغا أكثر سهولة في المستقبل إذا استمر الاقتصاد الرقمي بالنمو، وتزايد طلب على المزيد من الخدمات التي تقدم عبر الإنترنت من دون طرح أي أسئلة.

وكان أول هدف لهم هو موقع إم تورك MTurk، موقع أمازون ميكانيكال تورك Amazon Mechanical Turk، فبعد إطلاقه في عام 2005 كسوق تعهيدية جماعية على الإنترنت، نما بسرعة، وتنافس عشرات الآلاف من حول العالم -دون الإعلان عن هويتهم- على مدار الساعة لأداء مهام التداول التي تسمى المهام HITs، اختصاراً لعبارة Human Intelligence Tasks بمعنى: مهام الذكاء البشري. وتراوحت هذه المهام من تحويل التسجيلات الصوتية إلى نصوص كتابية، إلى تصنيف الصور، وكتابة وصف لصفحات الويب، وكانت كل هذه المهام مشتركة في أمر مشترك: إذا قمت بها بشكل جيد، لن يعرف أحد أنك ذكاء اصطناعي. وكان بمقدور بروميثيوس 10.0 أداء نحو نصف فئات المهام هذه بشكل جيد. ولكل فئة مهام من هذا القبيل، كان فريق أوميغا يجعل بروميثيوس يُطوِّر وحدة برمجيات Software module رشيقة مُصمَّمة خصيصاً، وهو عبارة عن وحدة برمجية ذكاء اصطناعي محدودة ومصممة خصيصاً لأداء هذه المهمة على وجه التحديد وليس سواها. وقد حمّلوا وحدة البرمجيات هذه على خدمات أمازون للويب Amazon Web Services، وهي منصة للحوسبة السحابية Cloud-computing platform بإمكانها تشغيل العديد من الأجهزة الافتراضية عند استئجارها. ومقابل كل دولار دفعوه إلى قطاع الحوسبة السحابية في أمازون، حصلوا على أكثر من دولارين من القطاع إم تورك في أمازون، ولم تكن شركة أمازون على علم بوجود مثل هذه الفرصة المذهلة للمراجحة داخل شركتهم!

ولتغطية آثارهم، أنشأ فريق أوميغا خلال الأشهر السابقة بخفاء الآلافَ من الحسابات إم تورك بأسماء أشخاص وهميين، و الآن انتحلت الوحدات التي طوّرها بروميثيوس هوياتهم. وعادة ما يدفع عملاء الحسابات إم تورك الرسوم بعد ثماني ساعات تقريباً، وعند هذه النقطة، يعيد فريق أوميغا استثمار الأموال في وقت أكثر للحوسبة السحابية، باستخدام وحدات مهام أفضل أنشأها أحدث إصدار من بروميثيوس المتطور باستمرار. ولأنهم تمكنوا من مضاعفة أموالهم كل ثماني ساعات؛ فإنهم سرعان ما بدؤوا يصلون إلى حد التشبع فيما يمكن تنفيذه من مهام الحسابات إم تورك، ووجدوا أنهم لا يستطيعون كسب أكثر من نحو مليون دولار في اليوم الواحد دون أن يثيروا انتباهاً غير مرغوب فيه بهم. ولكن هذا كان أكثر من كاف لتمويل خطوتهم التالية، والقضاء على أي حاجة لطلب الأموال من المدير المالي التنفيذي.

ألعاب خطرة

وبصرف النظر عن الإنجازات الخارقة التي حققوها في مجال الذكاء الاصطناعي، كان أحد المشاريع الأخيرة التي استمتع بها فريق أوميغا بمرح أكبر هو التخطيط لكسب الأموال بأسرع وقت ممكن بعد إطلاق بروميثيوس. وفي الأساس كان الاقتصاد الرقمي كله جاهزاً، ولكن هل من الأفضل البدء بتطوير ألعاب الحاسوب أم الموسيقى أم الأفلام أم البرمجيات

أم كتابة الكتب أم المقالات، أم التداول في سوق الأوراق المالية، أم تطوير اختراعات وبيعها؟ ويمكن اختصار هذا ببساطة إلى ما يمكّنهم من تحقيق أقصى معدل عائد على الاستثمار، ولكن استراتيجيات الاستثمار العادية كانت مجرد محاكاة ساخرة بطيئة لما يمكن أن يفعلوه: بينما قد يكون مستثمراً عادياً راضياً بعائد يبلغ 9% سنوياً، فإن استثماراتهم في الحسابات إم تورك حققت 9% في الساعة؛ مما يولد أموالاً أكثر بثمانية أضعاف كل يوم. والآن، بعد أن وصلوا إلى حد التَّشَبُّع في إم تورك، ماذا بعد؟

فكرتهم الأولى كانت تحقيق صفقة كبيرة في سوق الأوراق المالية - على أي حال، كان كل واحد من الفريق قد رفض في وقت ما عرضَ عملٍ مربحا من أجل تطوير الذكاء الاصطناعي لصناديق التحوط Hedge funds، التي كانت تستثمر بكثافة في هذه الفكرة بالضبط. وتذكر البعض أن الذكاء الاصطناعي حقّق أول مليون في فيلم **التسامي *Transcendence*** بهذه الطريقة. ولكن القيود الجديدة على المشتقات Derivatives بعد انهيار السوق في العام الماضي حدَّ من خياراتهم. وسرعان ما أدركوا أنه على الرغم من أنهم قد يحصلون على عوائد أفضل بكثير من المستثمرين الآخرين، فإنه من غير المحتمل أن يحصلوا على أي عائد يقترب مما يمكنهم الحصول عليه من بيع منتجاتهم الخاصة، عندما يكون لديك الذكاء الاصطناعي الفائق الأول في العالم، فالاستثمار في شركاتك الخاصة أفضل بكثير من الاستثمار في الآخرين! وعلى الرغم من أنه قد تكون هناك استثناءات عرضية (مثل استخدام قدرات بروميثيوس التي تفوق قدرة البشر في القرصنة Hacking على التغلغل في المعلومات الداخلية ثم شراء الأسهم التي هي على وشك الارتفاع)، فإن فريق أوميغا شعر بأن هذا لا يستحق الاهتمام غير المرغوب فيه الذي قد تثيره مثل هذه العمليات.

وعندما حولوا تركيزهم نحو المنتجات التي يمكنهم تطويرها وبيعها، بدت ألعاب الحاسوب أول خيار واضح. فسرعان ما قد يصبح بروميثيوس ذا مهارة فائقة في تصميم ألعاب جذابة، إذ يسهل عليه الترميز Coding، والتصميم الغرافيكي، وتتبع صور الأشعة، وجميع المهام الأخرى اللازمة لإنتاج منتج جاهز للشحن. إضافة إلى ذلك، بعد استيعاب كل بيانات الويب عن تفضيلات الأفراد، سيعرف بروميثيوس بالضبط ما يعجب كل فئة من اللاعبين، ويمكنه أن يطور قدرةً تفوق قدرةَ البشر على تحسين الألعاب بهدف زيادة أرباح المبيعات. فلعبة **صحائف الحكماء الإصدار الخامس: سكايريم *The Elder Scrolls V: Skyrim***، اللعبة التي أهدر فيها العديد من فريق أوميغا ساعات أكثر مما هم مستعدون للاعتراف بها، حققت أكثر من 400 مليون دولار خلال أول أسبوع في عام 2011، وكانوا واثقين من أن بروميثيوس قادر - خلال 24 ساعة- على أن يُطوِّر شيئاً يبعث على هذا القدر من الإدمان على الأقل باستخدام مليون دولار من موارد الحوسبة السحابية. ويمكنهم بعد ذلك بيعه عبر الإنترنت واستخدام بروميثيوس لانتحال شخصية البشر وهم يتحدثون عن اللعبة في عالم المدونات Blogosphere، إذا ولّد هذا 250 مليون دولار في أسبوع؛ فسيكونون قد ضاعفوا أرباحهم ثماني مرات في ثمانية أيام؛ مما يعطي عائداً بنسبة 3%

في الساعة - وهو أقل قليلا من بدايات إم تورك، ولكن أكثر استدامة. وبتطوير مجموعة من الألعاب الأخرى كل يوم، أدركوا أنهم سيكونون قادرين على كسب عشرة بلايين دولار سريعاً، ومن دون أن يصلوا إلى حدّ التَّشبُّع في سوق الألعاب.

لكن اختصاصية الأمن السيبراني في فريقهم أقنعتهم بالعدول عن خطة اللعبة هذه. فقد أشارت إلى أنها قد تشكل مخاطر غير مقبولة من **هروب** بروميثيوس والسيطرة على مستقبله بنفسه. وذلك لأنهم ليسوا متأكدين من كيفية تطوير بروميثيوس لأهدافه أثناء تحسينه الذاتي المطّرد، فقد قرر فريق أوميغا ألا يخاطروا وأن يبذلوا قصارى جهدهم للسيطرة على بروميثيوس («محاصرته في صندوق») بطرق تجعله لا يستطيع الهروب عبر الإنترنت. وبالنسبة إلى محرك بروميثيوس الرئيسي الذي يعمل في غرفة الخادم Server room، استخدموا قيوداً مادية: ببساطة لم يكن هناك اتصال بالإنترنت، وكان الناتج الوحيد من بروميثيوس في شكل رسائل ومستندات يُرسلها إلى حاسوب يسيطر عليه فريق أوميغا.

ومن ناحية أخرى، كان تشغيل أي برنامج معقد ابتكره بروميثيوس على حاسوب متصل بالإنترنت مقترحاً محفوفاً بالمخاطر: بما أن فريق أوميغا لم يكن لديه أي فهم كامل لما سيفعله، فلم تكن لديه أي وسيلة لمعرفة أنه لن يفعل ذلك. مثلاً، البدء بنشر نفسه عبر الإنترنت بسرعة كبيرة. وعند اختبار البرنامج الذي كتبه بروميثيوس لمهام إم تورك، حمى فريق أوميغا نفسه من ذلك عن طريق تشغيله فقط داخل جهاز افتراضي Virtual machine. وهذا عبارة عن برنامج يُحاكي حاسوباً: مثلاً، يشتري العديد من مستخدمي حواسيب ماك Mac برنامجاً افتراضياً يتيح لهم تشغيل البرامج ويندوز Windows عن طريق خداع الحاسوب إلى الاعتقاد أنه في الواقع في جهاز يعمل بالنظام ويندوز. وابتكر فريق أوميغا آلة افتراضية خاصة بهم أطلقوا عليها تحبباً اسم صندوق باندورا Pandora's Box، تُحاكي آلة مبسَّطة جدّاً أُزيلت منها جميع الإضافات التي ترتبط عادة بأجهزة الحاسوب: لا توجد لوحة مفاتيح، ولا شاشة، ولا مكبرات صوت، ولا اتصال بالإنترنت، ولا أي شيء آخر. فمثلاً عند تحويل تسجيلات المنصة إم تورك الصوتية إلى نصوص، حدَّد فريق أوميغا الأشياء بحيث إن جميع ما قد يدخل في صندوق باندورا كان مِلفاً صوتياً Audio file واحداً، وكل ما قد يخرج هو وثيقة نصية Text document واحدة - النص المكتوب. وقوانين الصندوق هذه كانت بالنسبة إلى البرمجيات الداخلية بمثابة قوانين الفيزياء لنا داخل عالمنا Univrse: لا يمكن للبرمجيات أن تسافر خارج الصندوق مثلما لا يمكننا السفر بأسرع من سرعة الضوء، بغض النظر عن مدى ذكائنا. وباستثناء هذه المدخلات Inputs والمخرجات Outputs، كان البرنامج داخل صندوق باندورا محصوراً بفعالية في عالَمٍ موازٍ ذي قواعد حوسبية Computational rules خاصة به. وكان لدى فريق أوميغا رهابٌ (فوبيا) عظيمٌ تخوفاً من هروب البرنامج إلى خارج الصندوق؛ مما دفعهم إلى إضافة صندوق يحيط بالزمن أيضاً، ويحدُّ من مدى العمر الافتراضي لشيفرة غير موثوق فيها. مثلاً، في كل مرة تنتهي فيها برمجية محصورة في صندوق من نسخ ملف صوتي واحد تُمسح أتوماتيكيّاً ذاكرة صندوق باندورا بالكامل ويعاد تثبيت البرمجيات من

جديد. وبهذه الطريقة، عندما يبدأ النظام بمهمة النسخ التالية، لا تكون لديه معرفة بما حدث من قبل، ومن ثم لا توجد قدرة على التعلم بمرور الوقت.

وعندما استخدم فريق أوميغا سحابةَ الأمازون لمشروعهم على المنصة إم تورك تمكنوا من أن يضعوا جميع وحدات مهامهم التي أنشأها بروميثيوس في مثل هذه الصناديق الافتراضية في السحابة؛ لأن مدخلات ومخرجات المنصة إم تورك كانت بسيطة جدّاً. ولكن هذا لن ينجح في ألعاب الحاسوب المثقلة برسومات الغرافيك، والتي لا يمكن وضعها في صناديق؛ لأنها تحتاج إلى الوصول التام إلى جميع أجهزة حاسوب اللاعب. وإضافة إلى ذلك، لم يرغبوا في المخاطرة بأن يقوم بعض المستخدمين المحنكين للحاسوب بتحليل رمز لعبتهم، واكتشاف صندوق باندورا، ويقرّر المستخدمين النظر فيما كان بداخله. ولا تقتصر مخاطر الهروب على سوق الألعاب فقط في الوقت الحالي، ولكن أيضاً على السوق المربحة بشكل كبير للبرمجيات الأخرى، السوق التي تقدر بمئات البلايين من الدولارات والمتاحة لمن يسعى إلى قطف ثمارها.

البلايين الأولى

كان فريق أوميغا قد حدّد نطاق بحثهم للوصول إلى منتجات ذات قيمة عالية، ورقمية بحتة (لتجنب التصنيع البطيء) ومفهومة بسهولة (مثلاً: النصوص أو الأفلام التي يعرفون أنها لن تكون معرّضة لخطر الهروب). وفي النهاية، قرروا إطلاق شركة إعلامية، بدءاً من الترفيه بالرسوم المتحركة، وكان كل من الموقع الإلكتروني وخطة التسويق والبيانات الصحفية جاهزة حتى قبل أن يصبح بروميثيوس فائق الذكاء - كل ما كان ينقص وقتها هو المحتوى.

وعلى الرغم من أن بروميثيوس كان مكتمل القدرات بشكل مذهل بحلول صباح يوم الأحد، يكدس النقود بانتظام من المنصة إم تورك، فإن قدراته الفكرية كانت لا تزال محدودة نوعاً ما: فقد تَمَّ تحسين بروميثيوس بشكل متعمد لتصميم أنظمة الذكاء الاصطناعي وكتابة البرمجيات التي تؤدي مهامَّ عبر المنصة إم تورك. ولكنه كان سيئاً مثلاً في صنع الأفلام، ليس لأي سبب عميق، ولكن للسبب نفسه الذي كان جيمس كاميرون James Cameron سيئاً في صنع الأفلام عند ولادته: هذه مهارة تتطلب وقتاً للتعلم. فمثل طفل بشري، بإمكان بروميثيوس أن يتعلم ما يريده من البيانات التي يمكنه الوصول إليها. وبينما استغرق جيمس كاميرون سنواتٍ ليتعلم القراءة والكتابة، فإن بروميثيوس تمكّن من فعل ذلك في يوم الجمعة، عندما وجد أيضاً وقتاً لقراءة كل من ويكيبيديا وبضعة ملايين من الكتب. ولكن صنع الأفلام كان أصعب، إن كتابة سيناريو مشوّق للبشر كان بصعوبة كتابة كتاب؛ مما تطلب فهماً مفصلاً للمجتمع الإنساني وما يَجِدُه البشر مسليا. وتحويل سيناريو إلى ملف فيديو نهائي يتطلب قدراً ضخماً من تعقب المسارات لمحاكاة الممثلين والمشاهد المعقدة التي يتحركون عبرها، ومحاكاة الأصوات، وإنتاج الموسيقى التصويرية الموسيقية المؤثِّرة وما إلى ذلك، اعتباراً من صباح يوم

الأحد، كان بإمكان بروميثيوس أن يشاهد فيلماً مدته ساعتان في نحو دقيقة واحدة، بما في ذلك أيضاً قراءة أي كتاب كان الفيلم يستند إليه وجميع المراجعات والتقييمات المنشورة على الإنترنت. ولاحظ فريق أوميغا أنه بعد أن شاهد بروميثيوس بضع مئات من الأفلام، صار أقدر على التنبؤ بنوع المراجعات التي سيحصل عليها فيلم ما، وما إذا كان سيروق القطاعات المختلفة من الجمهور. وبالتأكيد، تعلم بروميثيوس أن يكتب مراجعاته السينمائية الخاصة بطريقة تعكس بصيرة حقيقية، مراجعات تُعلِّق على كل شيء من الحبكات والتمثيل وصولاً إلى التفاصيل التقنية مثل الإضاءة وزوايا الكاميرا. ونظر فريق أوميغا إلى هذه القدرة على أنها تعني أنه عندما سيقوم بروميثيوس بعمل فيلمه الخاص؛ فإنه سيعرف ما يعنيه النجاح.

ووجَّه فريق أوميغا بروميثيوسَ إلى التركيز على إجراء الأبحاث في البداية، لتجنب الأسئلة المحرِجة حول من هم الممثلون الذين ستتم محاكاتهم. وفي ليلة الأحد، توّجوا عطلة نهاية الأسبوع الجامحة بتزويد أنفسهم بالمشروبات والفشار المُعد بالميكروويف، وتعتيم الأضواء ومشاهدة فيلم بروميثيوس الأول. وكان هذا الفيلم عبارة عن كوميديا خيالية بالرسوم المتحركة بروح فيلم ديزني «مُتجمِّد» ***Frozen***. وقد طُبِّقت تقنية تتبع الشعاع لتوليد الصور Ray tracking باستخدام شيفرة بروميثيوس المبنية في سحابة أمازون، وباستخدام معظم المليون دولار من الأرباح المتحققة على منصةإم تورك في يوم واحد. ومع بِدء الفيلم شعروا بالدهشة والخوف في الوقت نفسه من أن الفيلم صُنِع من قبل آلة دون توجيه إنسان، لكن سرعان ما كانوا يضحكون على النكات ويحبسون أنفاسهم خلال اللحظات الدرامية، بل إن البعض اغرورقت عيناه بعض الشيء عند النهاية العاطفية، منغمسين في هذا الواقع الخيالي لدرجة أنهم نسوا كل شيء عن صانعه.

وجَدْوَل فريق أوميغا إطلاق موقعهم ليوم الجمعة؛ مما يمنح بروميثيوس الوقتَ اللازمَ لإنتاج المزيد من المحتوى، وأنفسهم لأداء الأمور التي لا يستطيعون أن يعهدوا بها لبروميثيوس: شراء الإعلانات والبدء بتوظيف موظفين لشركات اسمية Shell أقاموها خلال الأشهر الماضية. ولإخفاء أثرهم، ستكون قصة التغطية الرسمية هي أن شركتهم الإعلامية (التي لم يكن لها أي ارتباط معلن بفريق بأوميغا) اشترت معظم محتواها من منتجي أفلام مستقلين، وهم عادة عبارة عن شركات ناشئة ذات تكنولوجيا عالية في المناطق ذات الدخل المنخفض، وهؤلاء الموردون المزيفون يعملون في مواقع نائية بشكل ملائم مثل تيروشيربالي Tiruch-chirappalli وياكوتسك Yakutsk، لا يعبأ معظم الصحافيين الفضوليين بزيارتها. أما الموظفون الوحيدون الذين وظفوهم بالفعل؛ فهم الذين يعملون في مجال التسويق والإدارة، وكانوا يخبرون أي شخص يطلب إجراء مقابلة أن فريق إنتاجهم يعمل في موقع مختلف ولا يجري مقابلات في الوقت الحالي. وبما يتلاءم مع قصتهم المزيفة، اختاروا شعار الشركة «توجيه موهبة العالم الإبداعية»، وصاغوا صورة الشركة على أنها مختلفة بشكل مُزعزع للسوق Disruptively، وذلك عن طريق استخدام أحدث التقنيات لتمكين المبدعين، وخاصة في العالم النامي.

وعندما حلّ يوم الجمعة وبدأ الزوار الفضوليون بالوصول إلى موقعهم، وجدوا شيئاً يذكرنا بخدمات الترفيه على الإنترنت مثل نيتفليكس Netflix وهولو Hulu ولكن باختلاف مثير للاهتمام، جميع مسلسلات الرسوم المتحركة كانت جديدة لم يسمعوا بها قط من قبل، وكانت آسرة جدّاً: معظم المسلسلات تألفت من 45 دقيقة من حلقات طويلة بحبكة قصصية مُحكمة، كل منها ينتهي بطريقة تجعلك متلهفاً إلى ما سيحدث في الحلقة القادمة. وكانت أرخص من المنافسين، حيث كانت أول حلقة من كل سلسلة مجانية، ويمكنك مشاهدة الحلقات الأخرى مقابل 49 سنتا لكل منها، مع تخفيضات على السلسلة الكاملة. وفي البداية، لم يكن هناك سوى ثلاثة مسلسلات بثلاث حلقات، ولكن تُضاف حلقات جديدة يوميّاً، إضافة إلى مسلسلات جديدة تستهدف الديموغرافيات المختلفة. وخلال الأسبوعين الأولين من بروميثيوس، تحسنت مهاراته في صناعة الأفلام بسرعة، ليس فقط من حيث الجودة، بل أيضاً من خلال خوارزميات أفضل لمحاكاة الشخصيات وتقنية الرسم بتتبع الشعاع؛ مما أدى إلى انخفاض كبير في تكلفة الحوسبة السحابية لصنع كل حلقة جديدة. ونتيجة لذلك، تمكن فريق أوميغا من إطلاق عشرات من المسلسلات الجديدة خلال الشهر الأول، واستهداف التركيبة السكانية بكاملها من الأطفال الصغار إلى البالغين، إضافة إلى التوسع في جميع أسواق اللغات الرئيسية في العالم؛ مما جعل موقعهم دوليّاً بشكل ملحوظ مقارنة بجميع المنافسين. وأبدى بعض المعلقين إعجابهم بحقيقة أنه ليس التسجيل الصوتي هو الذي كان متعدد اللغات، ولكن مقاطع الفيديو نفسها: مثلاً، عندما تتحدث شخصية إيطالية، تتطابق حركات الفم مع الكلمات الإيطالية، وكذلك إيماءات اليد الإيطالية المميزة. وعلى الرغم من أن بروميثيوس أصبح الآن قادراً تماماً على صنع الأفلام بمحاكاة ممثلين لا يمكن تمييزهم عن البشر، فقد تجنب فريق أوميغا هذا لعدم الكشف عن خطتهم. ولكن، أطلقوا العديد من المسلسلات بشخصيات متحركة شبه واقعية، وبتصنيفات فنية Genere تتناغم مع عروض الآكشن التلفزيونية الحية والأفلام التقليدية.

وسرعان ما صارت شبكتهم تبعث على الإدمان التام، وتمتعت بنمو مدهش في أعداد المشاهدين. ووجد العديد من المعجبين الشخصيات والحبكات أكثر ذكاء وأكثر إثارة للاهتمام حتى من أغلى أفلام الشاشة الضخمة في هوليود، وكانوا سعداء بأنهم يستطيعون مشاهدتها بتكلفة أقل بكثير. وقد انتعشت العوائد العالمية لتصل إلى عشرة ملايين دولار في اليوم خلال شهر من إطلاقها، وذلك بفضل الإعلانات الخارقة (التي يستطيع فريق أوميغا تحمل تكاليفها بفضل تكاليف إنتاجها التي تقارب الصفر)، والتغطية الإعلامية الممتازة، والعروض الجامحة والتوصيات الشخصية. وبعد شهرين تغلبوا على نيتفليكس، وبعد ثلاثة أشهر وصلوا إلى مرتبة دخل أعلى من مئة مليون دولار في اليوم، وبدؤوا ينافسون تايم وورنر Time Warner، وديزني Disney، وكومكاست Comcast وفوكس Fox، كواحدة من أكبر الإمبراطوريات الإعلامية في العالَم.

وحظي نجاحهم المثير بالكثير من الاهتمام غير المرغوب فيه، بما في ذلك التكهنات حول امتلاكهم لبرنامج ذكاء الاصطناعي قوي. ولكن باستخدام نسبة ضئيلة من إيراداتهم، أطلق فريق أوميغا حملة تضليل ناجحة إلى حد ما. ومن مكتب جديد في مانهاتن قام المتحدثون الرسميون الجدد بإضافة التفاصيل على قصصهم التي يتخفون وراءها، فتعاقدوا مع الكثير من الأشخاص كتمويه، بمن في ذلك كتاب سيناريو حقيقيون من جميع أنحاء العالم لبدء تطوير سلسلة جديدة، لا أحد منهم يعرف شيئاً عن بروميثيوس، فالشبكة الدولية المربكة للمقاولين من الباطن سهّلت على معظم موظفيهم افتراض أن آخرين في مكان آخر كانوا يؤدون معظم العمل.

لجعل أنفسهم أقل عرضة للخطر ولتجنب إثارة الانتباه بفعل الحوسبة السحابية المفرطة، وظّفوا أيضاً مهندسين للبدء ببناء سلسلة من الحواسيب الهائلة حول العالم، مملوكة من قبل شركات اسمية غير مرتبطة ببعضها ظاهرياً، وعلى الرغم من أنها مُسجَّلة مع السلطات المحلية على أنها «مراكز بيانات خضراء» Green Data Centers؛ لأنها كانت إلى حد كبير تعمل بالطاقة الشمسية، فإنها كانت حقيقة تركز بشكل رئيسي على الحوسبة وليس التخزين. وقد صممت بروميثيوس مخططاتها وصولاً إلى التفاصيل الأكثر دقة، باستخدام الأجهزة الجاهزة فقط، واختيارها لتقليل وقت البناء. ولم يكن لدى الأشخاص الذين بنوا هذه المراكز وأداروها أي فكرة عما تتم حوسبته في هذه الأجهزة: فقد ظنوا أنهم يديرون مرافق للحوسبة السحابية التجارية مماثلة لتلك التي تديرها أمازون وغوغل ومايكروسوفت، ويعرفون فقط أن جميع المبيعات كانت تُدار عن بعد.

تقنيات جديدة

وخلال أشهر بدأت إمبراطورية الأعمال التي يسيطر عليها فريق أوميغا في الحصول على موطئ قدم في مناطق أكثر فأكثر من الاقتصاد العالمي، بفضل التخطيط الفائق من قبل بروميثيوس. ومن خلال تحليل بيانات العالَم تحليلاً دقيقاً، فقد عرض بالفعل خلال الأسبوع الأول خطةَ النمو التفصيلية خطوة فخطوة، واستمر في تحسين وإعادة صياغة هذه الخطة مع نمو الموارد من البيانات والحواسيب. وعلى الرغم من أن بروميثيوس كان بعيداً كل البعد عن كونه يحيط علما بكل شيء، فإن قدراته أصبحت تفوق قدرات البشر بكثير، حتى إن فريق أوميغا نظر إليه على أنه الوسيط الروحاني Oracle المثالي، إذ يقدم بامتثال إجابات ونصائح رائعة ردّاً على جميع أسئلتهم.

وكانت برمجيات بروميثيوس قد صارت مثالية جدّاً للاستفادة إلى أقصى حدٍّ من الأجهزة المتوسطة الجودة التي اخترعها الإنسان والتي يعمل بها. وكما توقع فريق أوميغا، فقد حدّد بروميثيوس طرقاً لتحسين هذا الجهاز بشكل جذريّ. لكن خوفاً من هروبه، رفضوا بناء منشآت روبوتية يسيطر عليها بروميثيوس مباشرة. وبدلاً من ذلك، وظّفوا أعداداً كبيرة من العلماء والمهندسين العالميين في مواقع متعددة وزوّدوهم بتقارير أبحاث داخلية كتبها

بروميثيوس، متظاهرين بأنها من الباحثين في المواقع الأخرى. وقد كشفت هذه التقارير عن تأثيرات فيزيائية جديدة وتقنيات تصنيعية سرعان ما اختبرها مهندسوها وفهموها وأتقنوها. ودورات البحث والتطوير البشرية الطبيعية (R & D) بالطبع، تستغرق سنوات، وذلك إلى حد كبير لأنها تنطوي على العديد من الدورات البطيئة من التجربة والخطأ. ولكن الوضع الحالي كان مختلفاً تماماً: وبروميثيوس كان قد قام بالفعل بهذه الخطوات، لذلك كان العامل المحدد هو ببساطة مدى سرعة توجيه الناس نحو فهم الأشياء الصحيحة وبنائها. ويمكن للمدرس الجيد أن يساعد الطلبة على تعلم العلوم بشكل أسرع بكثير مما كان يمكن أن يكتشفوه من الصفر من تلقاء أنفسهم، وقد أدى بروميثيوس الدور نفسه مع هؤلاء الباحثين. وبما أن بروميثيوس يمكن أن يحدد بدقة المدة التي يستغرقها البشر لفهم وبناء الأشياء في ضوء الأدوات المختلفة، فقد طور أسرع مسار ممكن للتقدم قدما، مع إعطاء الأولوية للأدوات الجديدة التي يمكن فهمها وبناؤها بسرعة، والتي كانت بدورها مفيدة لتطوير أدوات أكثر تقدما.

وبروح حركة الصناع Maker movement، شُجِّعت الفرق الهندسية على استخدام أجهزتهم الخاصة لبناء آلاتهم الأفضل. ولم يُمكِّن هذا الاكتفاء الذاتي من الاقتصاد في المال فحسب، بل جعل الفريق أقل عرضة للتهديدات المستقبلية من العالم الخارجي. وفي غضون عامين كانوا ينتجون أجهزة حاسوب أفضل بكثير مما عرفه العالم؛ ولتجنب مساعدة المنافسة الخارجية، فقد أخفوا هذه التكنولوجيا واستخدموها فقط لتحسين بروميثيوس.

ولكن ما لاحظه العالم كان الطفرة التكنولوجية المذهلة. فقد بدأت الشركات الناشئة في جميع أنحاء العالم بإطلاق منتجات جديدة في كل المناطق تقريباً. وثمّة شركة ناشئة في كوريا الجنوبية أطلقت بطارية جديدة تخزن ضعف طاقة بطارية الحاسوب المحمول بنصف الكتلة، ويمكن شحنها في أقل من دقيقة. وأطلقت شركة فنلندية لوحة شمسية رخيصة بضعف كفاءة أفضل المنافسين. وأعلنت شركة ألمانية عن نوع جديد من الأسلاك فائقة التوصيل في درجة حرارة الغرفة، يمكن إنتاجها انتاجاً ضخماً؛ مما أحدث ثورة في قطاع الطاقة. وأعلنت مجموعة من شركات التكنولوجيا الحيوية في بوسطن عن تجربة إكلينيكية في المرحلة الثانية مما زعمت أنها لأول عقار فعّال وخالٍ من الآثار الجانبية لفقد الوزن، في حين أشارت الشائعات إلى أن شركة هندية كانت تبيع بالفعل شيئاً مماثلاً في السوق السوداء. وأجرت شركة في كاليفورنيا تجربة إكلينيكية في المرحلة الثانية على دواء مضاد للإصابة بالسرطان؛ مما جعل جهاز المناعة في الجسم يحدد ويهاجم الخلايا التي تحوي أياً من الطفرات السرطانية الأكثر شيوعاً. وظلت الأمثلة تتوارد تباعاً؛ مما أثار الحديث عن عصر ذهبي جديد للعلوم. وأخيراً وليس آخراً، كانت شركات الروبوتات تتزايد بسرعة في جميع أنحاء العالم، حتى وإن لم يقترب أي من البوتات من مستوى الذكاء البشري، ومعظمها لم يَبدُ شبيهاً بالبشر، لكنها زعزعت Disrupted الاقتصاد بشكل كبير. وعبر السنوات القادمة حلّت بشكل تدريجي محل معظم العمال في قطاعات التصنيع والنقل والتخزين والتجزئة والبناء والتعدين والزراعة وتحريج الغابات وصيد الأسماك.

وما لم يلاحظه العالم، بفضل العمل الشاق لفريق حاذق من المحامين، هو أن جميع هؤلاء كانوا يخضعون لفريق أوميغا من خلال سلسلة من الوسطاء. فقد كان بروميثيوس يُغرق مكاتب براءات الاختراع في العالم بسيل من براءات اختراع مثيرة عبر مختلف الوكلاء. وهذه الاختراعات أدت تدريجيّاً إلى الهيمنة في جميع مجالات التكنولوجيا.

وعلى الرغم من أن هذه الشركات الجديدة المُزَعزِعة خلقت أعداء أقوياء من منافسيها، فإنها كوّنت صداقات أكثر قوة. فقد كان فريق أوميغا رائداً بشكل استثنائي، وتحت شعارات مثل «الاستثمار في مجتمعنا»، أنفقوا قسماً كبيراً من هذه الأرباح في توظيف الموظفين المهرة في مشاريع مجتمعية - غالباً الأشخاص نفسهم الذين سرحتهم الشركات المُزَعزَعة. واستخدم الباحثون تحاليل مفصّلة أعدّها بروميثيوس تُحدِّد الوظائف التي ستكون مجزية إلى أقصى حدّ بالنسبة إلى الموظفين والمجتمع وبأقل تكلفة، ومصمّمة لتلائم الظروف المحلية. ففي المناطق التي تتمتع بمستويات عالية من الخدمات الحكومية، غالباً ما تركزت هذه الأعمال على تطوير المجتمع والأعمال الثقافية وتوفير الرعاية. أما في المناطق الفقيرة؛ فقد اشتملت أيضاً على بناء المدارس والرعاية الصحية والحضانة ورعاية المسنين والسكن بأسعار معقولة والحدائق والبنية التحتية الأساسية. وإلى حد كبير في كل مكان؛ فقد اتفق السكان المحليون على أن هذه هي الأمور التي كان ينبغي فعلها منذ فترة طويلة. وحصل السياسيون المحليون على تبرعات سخية، واتُخِذ اللازم لجعلهم يبدون في صورة حسنة ليشجعوا هذه الاستثمارات المجتمعية من قبل الشركات.

كسب السلطة

وكان فريق أوميغا قد أطلق شركة إعلامية ليس فقط للترويج لمشاريعها التقنية المبكرة، ولكن أيضاً للخطوة التالية من خطتها الجريئة: الاستيلاء على العالم. وفي غضون عام من الإطلاق الأول، أضافوا إلى تشكيلتهم في جميع أنحاء العالم قنواتٍ إخباريةً جيدةً بشكل لافت للنظر. وعلى عكس قنواتهم الأخرى، فقد تَمَّ تصميم هذه القنوات عمداً لخسارة المال، وزُوِّج لها على أنها خدمة عامة. وفي الواقع، لم تولّد قنواتهم الإخبارية أي دخل على الإطلاق: فلم تعرض إعلانات، وكانت متاحة من دون رسوم لأي شخص متصل بالإنترنت. وقد كانت بقية إمبراطوريتهم الإعلامية آلة توليد نقد تولِّد نقدا كافيا يمكِّنَهم من إسباغ الموارد على خدمتهم الإخبارية بما هو أكثر بكثير من أي جهد صحفي آخر في تاريخ العالم. وتمثل ذلك بالسعي المحتدم إلى التوظيف بفاعلية والرواتب التنافسية المقدمة للصحفيين والمراسلين الإخباريين، فجلبوا مهارات واكتشافات رائعة إلى الشاشة. ومن خلال خدمة ويب عالمية تدفع لأي شخص يكشف عن شيء يستحق النشر، من حدث فساد محلي إلى القصص المؤثرة عاطفيّاً، صاروا في العادة الأوائل في تحقيق السبق الصحفي. على الأقل هذا ما اعتقده الأفراد: في الواقع، كانوا في الغالب الأوائل؛ لأن بروميثيوس كان يكتشف القصص المنسوبة إلى الصحفيين من المواطنين عبر المراقبة الفورية للإنترنت، كل هذه المواقع الإخبارية بالفيديو وفَّرت خدمة بودكاست وكذلك المقالات المطبوعة.

كانت المرحلة الأولى من استراتيجيتهم الإخبارية آخذة بكسب ثقة الناس، وقد حققوا نجاحاً كبيراً، مكنهم استعدادهم غير المسبوق لخسارة المال من توفير تغطية إخبارية إقليمية ومحلية مثابرة بشكل ملحوظ. وفي كثير من الأحيان كشف المراسلون الميدانيون عن فضائح تستحوذ حقاً على اهتمام مشاهديها. وحينما يكون بلدٌ ما منقسماً سياسياً ومعتاداً على الأخبار الحزبية، فإن فريق أوميغا سيطلق قناة إخبارية لكل فصيل، تمتلكها ظاهرياً شركات مختلفة، وتكسب ثقة ذلك الفصيل تدريجيّاً. وحيثما أمكن، حققوا ذلك باستخدام الوكلاء لشراء القنوات القائمة الأكثر تأثيراً، وتحسينها تدريجيّاً بالتخلص من الإعلانات وإدخال محتواهم الخاص. أما في البلدان التي تهدد فيها الرقابة والتدخلات السياسية هذه الجهود؛ فإنهم في البداية يقبلون بجميع اشتراطات الحكومة ليظلوا عاملين فيها، مع شعار داخلي سري «الحقيقة، لا شيء غير الحقيقة، ولكن ربما ليس الحقيقة كاملة». ويقدم بروميثيوس عادة نصيحة ممتازة في مثل هذه المواقف، فيوضح أي السياسيين يجب تسليط الأضواء الإيجابية عليه وأيُّهم يجب أن يُفضح (عادة هم الفاسدون المحليون)، كما قدم بروميثيوس توصيات لا تقدر بثمن لمعرفة الخيوط التي يجب سحبها، ومن تجب رشوته، وأفضل طريقة لفعل بذلك.

حققت هذه الاستراتيجية نجاحاً باهراً في جميع أنحاء العالم، فبرزت القنوات الناشئة التي يسيطر عليها فريق أوميغا على أنها مصادر الأخبار الأكثر موثوقية. وحتى في البلدان التي كانت الحكومات قد أحبِطت فيها محاولات انتشارهم الواسع، فإنهم بنوا سمعة جديرة بالثقة، وكثير من قصصهم الإخبارية كانت تتسرب عبر وسائل الاتصال غير الرسمية أو الإشاعات. أما الإخباريون المنافسون؛ فقد شعروا بأنهم كانوا يخوضون معركة ميؤوساً منها: كيف يمكن أن تتنافس مع شخص يتمتع بتمويل أفضل ويوفر منتجاته مجاناً؟ مع انخفاض نسبة المشاهدة، قررت المزيد من الشبكات بيع قنواتها الإخبارية إلى مجموعة شركات اتضح لاحقاً أنها تابعة لفريق أوميغا.

وبعد نحو عامين من إطلاق بروميثيوس، عندما اكتملت مرحلة اكتساب الثقة إلى حد كبير، أطلق فريق أوميغا المرحلة الثانية من استراتيجيته الإخبارية: الإقناع. وحتى قبل ذلك كان المراقبون الأذكياء قد لاحظوا إشاراتٍ تلمح إلى أجندة سياسية متوارية وراء الإعلام الجديد: يبدو أن هناك دفعة لطيفة تجاه الوسط، بعيداً عن أي تطرف من أي نوع. فمجموعة القنوات العديدة التي تروج لاحتياجات المجموعات المختلفة لا تزال تعكس العداوة بين الولايات المتحدة وروسيا، والهند وباكستان، والأديان المختلفة والفصائل السياسية وما إلى ذلك. ولكن الانتقادات انخفضت قليلاً، وعادة ما تركز على قضايا محددة تتعلق بالمال والسلطة بدلاً من على تجريح الشخصيات، والأخبار المثيرة للذعر والشائعات غير المُتحقّق منها جيداً. بمجرد بدء المرحلة الثانية صارت هذه الدفعة لنزع فتيل الصراعات القديمة أكثر وضوحاً، مع قصص مؤثرة متكررة حول محنة الأعداء التقليديين المختلطة مع تقارير إخبارية حول كون مثيري النزاع جَهوريي الصوت مدفوعين بدوافع شخصية.

أشار المعلقون السياسيون إلى أنه بالتوازي مع التقليل من التركيز على النزاعات الإقليمية، بدا أن هناك دفعة متضافرة نحو الحد من التهديدات العالمية. مثلاً، فجأة صارت

مخاطر الحرب النووية تُناقش في كل مكان، فظهر العديد من الأفلام الرائجة بسيناريوهات تبدأ فيها الحرب النووية العالمية مصادفة أو عن قصد، وبالغت في عرض مآسي الشتاء النووي، وانهيار البنية التحتية والمجاعة العارمة. وأوضحت الأفلام الوثائقية الجديدة كيف أن الشتاء النووي يمكن أن يؤثر في كل بلد. ومُنِح العلماء والسياسيون الذين يدافعون عن تخفيف التصعيد النووي وقتاً طويلاً على الشاشات، ليس أقلها الوقت المخصص لمناقشة نتائج العديد من الدراسات الجديدة حول التدابير المفيدة التي يمكن اتخاذها - الدراسات الممولة من قبل المؤسسات العلمية التي تلقت تبرعات كبيرة من شركات التكنولوجيا الجديدة. ونتيجة لذلك، بدأ الزخم السياسي بالتزايد مطالباً بخفض مستوى تأهب الصواريخ وتقليص الترسانات النووية، وجه الاهتمام المتجدد بالإعلام إلى التغير العالمي في المناخ، مُبرِزاً في كثير من الأحيان الاختراقات التكنولوجية الحديثة التي صارت ممكنة بفعل بروميثيوس، والتي حفضت تكلفة الطاقة المتجددة؛ مما شجع الحكومات على الاستثمار في مثل هذه البنية التحتية الجديدة للطاقة.

وبالتوازي مع سيطرتهم على وسائل الإعلام، سَخَّر فريق أوميغا بروميثيوس مواردهم لإحداث ثورة في التعليم. فبالنظر إلى معرفة أي شخص وقدراته، يمكن لبروميثيوس تحديد أسرع طريقة لتعلُّمهم أي موضوع جديد بطريقة تجعلهم مُتَفاعلين ومُحفَّزين على الاستمرار، وينتج روميثيوس مقاطعَ الفيديو المرئية الضرورية ومواد القراءة والتمارين وأدوات التعلّم الأخرى. ولذلك، سوَّقت الشركات التي يسيطر عليها فريق أوميغا الدورات عبر الإنترنت في كل موضوع تقريباً، وهي مصمَّمة بشكل كبير ليس فقط لتلائم اللغة والخلفية الثقافية ولكن أيضاً مستوى البدء. سواء أكنت أُميّاً في الرابعة والأربعين من عمرك وترغب في تعلم القراءة أم عالِم أحياء بدرجة دكتوراه تبحث عن تطوير أحدث علاج مناعي للأورام، فإن بروميثيوس لديه الدورة المثالية لك. وهذه العروض لا تشبه الدورات التدريبية على الإنترنت في الوقت الحاضر إلا شبهاً بسيطاً: من خلال الاستفادة من مواهب بروميثيوس في صنع الأفلام، فإن مقاطع الفيديو ستستقطب حقّاً مشاركة المتعلم، فالاستعارات القوية التي قد تدركها ستولد لديك شغفاً لمعرفة المزيد. وقُدِّمت بعض الدورات التدريبية برسوم بهدف الربح، لكن العديد توفَّر مجاناً. وهو أمر أسعد المُعلمين في جميع أنحاء العالم الذين تمكنوا من استخدامها في فصولهم الدراسية، وكذلك معظم الأشخاص الذين يتوقون إلى تعلم أي شيء.

وأثبتت هذه القوى العظمى التعليمية أنها أدوات فعالة لأغراض سياسية؛ مما أدى إلى إنشاء «متتاليات الإقناع» Persuasion sequences عبر الإنترنت باستخدام مقاطع الفيديو التي تؤدي فيها التبصرات من كلٍّ منها إلى تحديث وجهات نظر أحد الأشخاص وتحفِّزه مشاهدة فيديو آخر حول الموضوع ذي الصلة؛ مما يرجح تعزيز اقتناعهم أكثر فأكثر. فمثلاً، عندما كان الهدف هو نزع فتيل نزاع بين دولتين، فستوزّع الوثائق التاريخية -بشكل مستقل في كل من البلدين- التي تسلط الضوء على أصل النزاع فتعالجه من زوايا أكثر دقة، ستشرح القصص الأخبارية التعليمية من هم المستفيدون من استمرار النزاع وأساليبهم في إذكائه، وفي الوقت نفسه، ستبدأ الشخصيات المحبوبة من الأمة الأخرى

بالظهور في العروض الشعبية على القنوات الترفيهية، تماماً مثلماعزز تصوير شخصيات الأقليات بطريقة متعاطفة حركات الحقوق المدنية في الماضي.

وقبل مُضِّي فترة طويلة، لم يعد بإمكان المعلقين السياسيين إلا أن يلاحظوا تزايد الدعم لأجندة سياسية تتركز حول سبعة شعارات:

1. الديمقراطية
2. تخفيضات ضريبية
3. تخفيضات الخدمات الاجتماعية الحكومية
4. تخفيضات الإنفاق العسكري
5. التجارة الحرة
6. فتح الحدود
7. الشركات المسؤولة اجتماعيّاً

غير أن ما ظلَّ أقل وضوحاً كان هو الهدف الأساسي: تقويض جميع هياكل السلطة السابقة في العالم. أدت البنود 2 إلى 6 إلى تآكل سلطة الدولة. ومنحت دمقرطةُ العالم إمبراطورية أوميغا التجارية مزيداً من التأثير في اختيار القادة السياسيين. إذ أدت الشركات المسؤولة اجتماعيّاً إلى إضعاف سلطة الدولة من خلال الاستحواذ على المزيد من الخدمات التي كانت الحكومات تُقدِّمها (أو ينبغي أن تقدمها).

فقد ضعفت نخبة رجال الأعمال التقليديين ببساطة؛ لأنها غير قادرة على منافسة الشركات المدعومة من قبل بروميثيوس في السوق الحرة، ومن ثم كانت حصتها من الاقتصاد العالمي آخذة بالتقلص. وافتقر قادة الرأي التقليديون، من الأحزاب السياسية إلى الجماعات الدينية، إلى آلية إقناع قادرة على منافسة الإمبراطورية الإعلامية لأوميغا.

كما هي الحال مع أي تغيير شامل، كان هناك رابحون وخاسرون. وعلى الرغم من نشوء شعور جديد بالتفاؤل في معظم البلدان مع تحسن التعليم والخدمات الاجتماعية والبنية التحتية، فقد تراجعت الخلافات، وأطلقت الشركات المحلية تكنولوجيات هائلة اكتسحت العالم، غير أن الجميع لم يكونوا سعداء. ففي الوقت الذي حصل فيه العديد من العمال المُسرَّحين على إعادة توظيفهم في مشاريع مجتمعية، فإن أولئك الذين كانوا يمتلكون قوة وثروة كبيرتين رأوهما تتناقصان بشكل عام، بدأ هذا في قطاعي الإعلام والتكنولوجيا، لكنه انتشر في كل مكان تقريباً. وأدى انخفاض النزاعات العالمية إلى تخفيضات ميزانية الدفاع التي أضرت بمتعهدي الخدمات العسكرية، وعادة لم تكن الشركات الناشئة المزدهرة متداولة في سوق الأسهم، مبرِّرين ذلك بأن حملة الأسهم الذين يرغبون في زيادة الأرباح إلى الحد الأقصى سيحُول ذلك دون إنفاقهم الضخم على المشاريع المجتمعية. وهكذا ظلت أسواق الأسهم العالمية تفقد قيمتها، مهددة كلّاً من أقطاب الأعمال أو المواطنين العاديين الذين يعتمدون على صناديق معاشاتهم التقاعدية. وإذا لم يكن تقلص أرباح الشركات المتداولة علناً سيئاً بما فيه الكفاية، فقد لاحظت استثمارات المستثمرين في

جميع أنحاء العالم توجهاً مزعجاً: جميع خوارزميات التداول التي كانت ناجحة في السابق على ما يبدو لم تعد صالحة، وضعف أداء حتى مؤشرات الصناديق البسيطة، وبدا أن هناك شخصاً آخر يتفوَّق عليهم دائماً ويهزمهم في لعبتهم.

وعلى الرغم من أن الجماهير الشعبية قاومت موجة التغيير، فإن ردة فعلها كانت غير فعالة بتاتاً، كما لو كانوا قد وقعوا في فخ مُخطَّط له بشكل جيد. وكانت التغيرات الهائلة تحدث بوتيرة مذهلة لدرجة يصعب تتبعها وتطوير استجابة منسقة ضدها. إضافة إلى ذلك، فقد كان من غير الواضح إلى حد كبير ما الذي ينبغي أن يُطالب به، كان اليمين السياسي التقليدي قد شهد تحقيق معظم شعاراته. ولكن التخفيضات الضريبية وتحسين مناخ الأعمال كانا يساعدان في الغالب منافسيهم من ذوي القدرات التكنولوجية الأعلى، وفي الواقع، كانت كل الصناعات التقليدية تطالب الآن بمن يُنقذها، لكن الصناديق الحكومية المحدودة بالكاد ساعدتها في معركة يائسة ضد بعضها البعض، بينما صَوَّرتهم وسائل الإعلام على أنهم ديناصورات تسعى إلى الحصول عى دعم الدولة لمجرد أنها لا تستطيع المنافسة. وعارض اليسارُ السياسيُّ التقليديُّ التجارة الحرة والتخفيضات في الخدمات الاجتماعية الحكومية، لكنه سعد بالتخفيضات العسكرية والحد من الفقر. وبالفعل، تضعضع الكثير من وعيدهم بفعل الحقيقة التي لا يمكن إنكارها والتي مفادها أن الخدمات الاجتماعية تحسنت الآن بعد أن صارت توفرها الشركات المثالية بدلاً من الدولة. وأظهر استطلاع رأي بعد استطلاع رأي آخر أن معظم الناخبين في جميع أنحاء العالم يشعرون بتحسن نوعية حياتهم، وأن الأمور تسير بشكل عام في اتجاه جيد. وكان لهذا تفسير رياضياتي بسيط: قبل بروميثيوس، كان أفقر 50% من سكان الأرض يحصلون على نحو 4% فقط من الدخل العالمي؛ مما مكّن الشركات التي تسيطر عليها أوميغا من كسب قلوبهم (والأصوات) من خلال تقاسمهم جزءاً بسيطاً فقط من أرباحهم.

التوطيد

ونتيجة لذلك، شهدت الأمة انتصارات ساحقة هائلة للأحزاب التي تبنَّت شعارات أوميغا السبعة، ففي الحملات المُحسَّنة بعناية، صوروا أنفسهم على أنهم يقعون في منتصف الطيف السياسي، وشجبوا اليمين كمتخلفين جشعين يطالبون بإنقاذ الشركات، وألقوا باللوم على اليسار كمعاندين للابتكار في مجال الضرائب والصرف الحكوميين. وما لم يكن أحد يدركه تقريباً هو أن بروميثيوس اختار بعناية أفضل الأفراد ليُعدّوا المُرشحين، واستخدم جميع نفوذه لتأمين انتصارهم.

وقبل بروميثيوس كان هناك دعم متنامٍ لدخل أساسي متساوٍ، اقترح حدّاً أدنى ممولاً من الضرائب كعلاج للبطالة التكنولوجية. وانهارت هذه الحركة عندما انطلقت المشاريع المجتمعية للشركات؛ لأن إمبراطورية الأعمال التي تسيطر عليها أوميغا كانت في الواقع تقدم الشيء نفسه. وبذريعة تحسين تنسيق مشاريعهم المجتمعية، أطلقت مجموعة

دولية من الشركات التحالف الإنساني Humanitarian Alliance، والتحالف منظمة غير حكومية تهدف إلى تحديد أهم الجهود الإنسانية وتمويلها في جميع أنحاء العالم. وسرعان ما ساندتها إمبراطورية أوميغا بالكامل تقريباً، وأطلقت مشاريع عالمية على نطاق غير مسبوق، حتى في البلدان التي تخلفت إلى حد كبير عن اللحاق بطفرة التكنولوجيا، وتحسين التعليم والصحة والازدهار والحوكمة. وبالطبع، قَدَّم بروميثيوس خطط مشاريع مدروسة بعناية وراء الكواليس، ومرتبة حسب التأثير الإيجابي لكل دولار يُصرف. فبدلاً من مجرد صرف الأموال، كما هي الحال في مقترحات الدخل الأساسي، فإن التحالف (كما صار معروفاً بالعامية) سيشرك أولئك الذين دعمهم ليعملوا من أجل قضيته. ونتيجة لذلك، في النهاية انتهى الأمر بجزء كبير سكان العالم إلى الشعور بالامتنان والولاء للتحالف - وغالباً ما كان ذلك أكثر من امتنانهم لحكوماتهم.

ومع مرور الوقت، فقد اتخذ التحالف على نحو متزايد دور حكومة عالمية، فشهدت الحكومات الوطنية تآكل قوتها باستمرار، وظلت الميزانيات الوطنية تتقلص بسبب التخفيضات الضريبية في حين نمت ميزانية التحالف وتقزمت أمامها ميزانيات جميع الحكومات مجتمعة. وتدريجيّاً تزايد تحول جميع الأدوار التقليدية للحكومات الوطنية إلى أدوار مكرّرة لا حاجة إليها وغير ذات صلة. فقد وفّر التحالف أفضل الخدمات الاجتماعية والتعليم والبنية التحتية، ونجحت وسائل الإعلام في نزع فتيل النزاع الدولي لدرجة أن الإنفاق العسكري صار غير ضروري إلى حد كبير، ومحا الرخاء المتنامي معظم جذور الصراعات القديمة، التي ترجع إلى التنافس على الموارد الشحيحة. وعدد قليل من الدكتاتوريين وغيرهم قاوم بعنف هذا العالم الجديد ورفض أن يُشترى، لكنهم جميعاً سقطوا في انقلابات مُدبَّرة بعناية أو انتفاضات شعبية.

وهكذا استكمل فريق أوميغا التحول الأكثر دراماتيكية في تاريخ الحياة على الأرض، للمرة الأولى على الإطلاق، كان كوكبنا تُديره سلطة واحدة، مُضخَّمة بواسطة ذكاء فائق لدرجة أنه كان من الممكن أن يُمكّن الحياة من أن تزدهر لملايين السنين على الأرض وعبر الكوزموس، ولكن ما هي خططهم تحديداً؟

***

كانت تلك قصة فريق فريق أوميغا، غير أن بقية هذا الكتاب تدور حول قصة أخرى - قصة لم تكتب بعد: حكاية مستقبلنا مع الذكاء الاصطناعي، كيف تريد لذلك أن يتكشّف؟ هل يمكن أن يحدث شيء ما يشبه ولو قليلاً قصة مثل قصة فريق أوميغا، وإذا كان الأمر كذلك، فهل تريد ذلك؟ وإذا تركنا جانباً التخمينات حول الذكاء الاصطناعي، فكيف تريد أن تبدأ حكايتنا؟ كيف تريد أن يؤثر الذكاء الاصطناعي في الوظائف والقوانين والأسلحة بالعقد القادم؟ ناظراً إلى الأمام، كيف تكتب النهاية؟ فهذه الحكاية هي ذات أبعاد كوزموسية حقيقية؛ لأنها تنطوي على مستقبل الحياة الأقصى في عالمنا، إنها قصة متاح لنا أن نكتبها.

الفصل 1

مرحباً بك في أهم محادثة في زمننا

«التكنولوجيا تمنحُ الحياةَ القدرةَ على أن تزدهر أو أن تدمر نفسها ذاتيّاً بمثل ما لم يسبق لهما من مثيل».

معهد مستقبل الحياة

بعد 13.8 بليون سنة من ولادته، استيقظ كوننا Universe وأصبح مدركاً لذاته. ومن كوكب أزرق صغير، بدأت أجزاء واعية صغيرة من كوننا تحدق في الكوزموس Cosmos بالتلسكوبات، وتكتشف مراراً وتكراراً أن كل ما يُعتقد أنه موجود هو مجرد جزء صغير من شيء أعظم: نظام شمسي، ومجرة، وكون بأكثر من مئة بليون مجرة أخرى مرتبة في نمط متقن من المجموعات والعناقيد والعناقيد الفائقة. وعلى الرغم من اختلاف مراقبي النجوم هؤلاء أنفسهم حول العديد من الأمور، فإنهم يميلون إلى الاتفاق على أن هذه المجرات جميلة ومُلهِمة.

ولكن الجمال في نظر الناظر، وليس في قوانين الفيزياء، لذا قبل أن يستيقظ كوننا، لم يكن هناك جمال. وهذا يجعل صحوتنا الكوزموسية أكثر روعة وجديرة بالاحتفال: فقد حولت هذه الصحوة كوننا من زومبي لا عقل له وبلا وعيٍّ بذاته إلى نظام إيكولوجي حيّ يؤوي التأملَ الذاتيَّ والجمال والأمل - والسعي إلى تحقيق الأهداف والمعنى والغرض. ولو لم يستيقظ كوننا قطُّ، فعندئذ بقدر ما يعنيني؛ لكان غير مُجْدٍ على الإطلاق ومجرد فراغ هائل من الفضاء. وإذا عاد كوننا إلى السبات بشكل دائم بسبب بعض كارثة كوزموسية أو كارثة ذاتية نتسبب بها إلى أنفسنا؛ فسيصبح، للأسف، بلا معنى.

من ناحية أخرى، قد تتحسن الأمور، نحن لا نعرف حتى الآن ما إذا كنا نحن البشر مراقبي النجوم الوحيدين في كوكبنا، أو حتى أولهم، لكننا تعلمنا بالفعل ما يكفي عن كوننا لنعرف أن لديه القدرة على الاستيقاظ بشكل أفضل بكثير مما حصل حتى الآن. ربما نشبه بصيصاً خافتاً من الوعي الذاتي الذي تستشعره عندما استيقظت من النوم هذا الصباح: هاجس من وعي أعظم

سيتحصل بمجرد فتح عينيك واستيقاظك تماماً. وربما ستنتشر الحياة في كل أرجاء الكوزموس وتزدهر لبلايين أو تريليونات من السنين - وربما يكون ذلك بسبب القرارات التي نتخذها هنا على كوكبنا الصغير خلال حياتنا.

تاريخ موجز للتعقيد

إذاً، كيف حدث هذا الاستيقاظ المذهل؟ لم يكن هذا حدثاً معزولاً، بل كان مجرد خطوة في عملية مستمرة بلا هوادة لمدة تبلغ 13.8 بليون عام جعلت من كوننا أكثر تعقيداً وإثارة للاهتمام - وهذه عملية مستمرة بوتيرة متسارعة.

بصفتي عالِم فيزياء، أشعر بأنني محظوظ لأنني قضيت الغالب الأعم من ربع القرن الماضي للمساعدة على تحديد تاريخنا الكوزموسي، وكانت رحلة مذهلة من الاكتشاف. فمنذ الأيام التي كنت فيها طالب دراسات عليا، انتقلنا عن الجدل حول ما إذا كان عمرنا يبلغ من 10 أو 20 بليون عام لنتجادل حول ما إذا كان عمره 13.7 أو 13.8 بليون سنة، وذلك بفضل مجموعة من التلسكوبات الأفضل، وأجهزة الحاسوب الأفضل والفهم الأفضل. ولا يزال علماء الفيزياء لا يعرفون على وجه اليقين ما الذي تسبب بالانفجار الكبير (العظيم) Big bang، أو ما إذا كان هذا بالفعل بداية كل شيء أو مجرد تتمة لمرحلة مبكرة. ولكننا اكتسبنا فهماً تفصيلياً إلى حدٍ ما لما حدث منذ الانفجار الكبير، وذلك بفضل التراكمات الضخمة المتتالية من المقاييس عالية الجودة، لذا يرجى إتاحة بضع دقائق لي لتلخيص 13.8 بليون سنة من تاريخ الكوزموس.

في البدء كان الضوء، ففي أول جزء من الثانية بعد الانفجار الكبير، كان الجزء الكامل من الفضاء، الذي يمكن أن ترصده تلسكوباتنا ("الكون الذي يمكن رصده" أو ببساطة "كوننا" للاختصار) أكثر سخونة وأكثر سطوعاً من لب شمسنا، وتوسع بسرعة. وعلى الرغم من أن هذا قد يبدو مذهلاً، فإنه كان أيضاً مملاً بمعنى أن كوننا لم يكن يحتوي إلا على حساء متجانس وكثيف وساخن ومضجر من الجسيمات الأولية. وبدت الأشياء متشابهة إلى حد كبير في كل مكان، وكانت البُنية الوحيدة المثيرة للاهتمام تتألف من موجات صوتية خافتة ذات مظهر عشوائي جعلت بعض أجزاء الحساء أكثر كثافة بنحو 0.001%. ويسود الاعتقاد أن هذه الموجات الخافتة نشأت عما يطلق عليه التقلبات الكمّيّة Quantum fluctuations؛ لأن مبدأ الشك Uncertainty principle لهايزنبرغ في ميكانيكا الكم Quantum mechanics يمنع أي شيء من أن يكون مُمِلاً وموحَّداً تماماً.

ومع توسع كوننا وفقده للحرارة، صار أكثر إثارة للاهتمام مع اندماج جزيئاته في أجسام أكثر تعقيدا، خلال أول جزء من الثانية، جمعت القوة النووية القوية الكواركات معا مُشكِّلة بروتونات (نواة الهيدروجين) ونيوترونات، واندمج بعضها بدوره في نوى الهيليوم بغضون بضع دقائق. وبعد نحو 400 ألف سنة، ربطت القوة الكهرومغناطيسية Electromagnetic force بين هذه النوى والإلكترونات لتصنع الذرات الأولى. ومع استمرار تمدد كوننا، بردت هذه الذرات تدريجيّاً إلى غاز مُعتم وبارد، واستمرت هذه الليلة الأولى لمدة 100 مليون عام تقريباً. وهذه الليلة الطويلة قادت إلى فجرنا الكوزموسي عندما نجحت قوة الجاذبية في تضخيم تلك التقلبات

في الغاز، وسحب الذرات معاً لتشكيل النجوم الأولى والمجرات. وهذه النجوم الأولى ولَّدت الحرارة والضوء عن طريق دمج الهيدروجين في ذرات أثقل مثل الكربون والأكسجين والسيليكون. وعندما ماتت هذه النجوم، أُعيد تدوير العديد من الذرات التي صنعتها في الكوزموس؛ فشكَّلت كواكب حول نجوم الجيل الثاني.

وفي مرحلة ما، ترتبت مجموعة من الذرات في نمط معقد قد يحافظ على نفسه ويكرر نفسه. وسرعان ما كانت هناك نسختان، وظل العدد يتضاعف. ولا يتطلب الأمر سوى 40 تكراراً للوصول إلى تريليون، لذا سرعان ما صار هذا المُضاعِف الذاتي الأول قوة لا يستهان بها. وانبثقت الحياة إلى الوجود.

المراحل الثلاث للحياة

والسؤال حول كيفية تعريف الحياة هو أمر مثير للجدل، فهناك العديد من التفسيرات المتنافسة، بعضها يتضمن متطلبات محددة جدّاً مثل كونها مكونة من خلايا؛ مما قد يؤدي إلى استبعاد كل من الآلات الذكية المستقبلية والحضارات خارج كوكب الأرض. ولمّا كنا لا نريد الحد من تفكيرنا في شأن مستقبل الحياة للأنواع Species التي واجهناها حتى الآن، فدعونا بدلاً من ذلك نُعرِّف الحياة تعريفا واسع النطاق، أي ببساطة كعملية قد تحافظ على تعقيدها وتكرار نفسها. وما يتم نسخه ليس مادة (مصنوعة من ذرات) بل معلوماتٍ (مصنوعة من البتات Bits) تُحدد كيفية ترتيب الذرات. وعندما تصنع البكتيريا نسخةً من حمضها النووي DNA، لا تُنشأ ذرات جديدة، ولكن تُرتَّب مجموعة جديدة من الذرات على النمط الأصلي نفسه، ومن ثم فهي تنسخ المعلومات. وبعبارة أخرى، يمكننا أن نفكر في الحياة كنظام لمعالجة المعلومات ينسخ نفسه ذاتيّاً، والذي تُحدد معلوماته (برمجياته) سلوكَه ومخططاته الخاصة به.

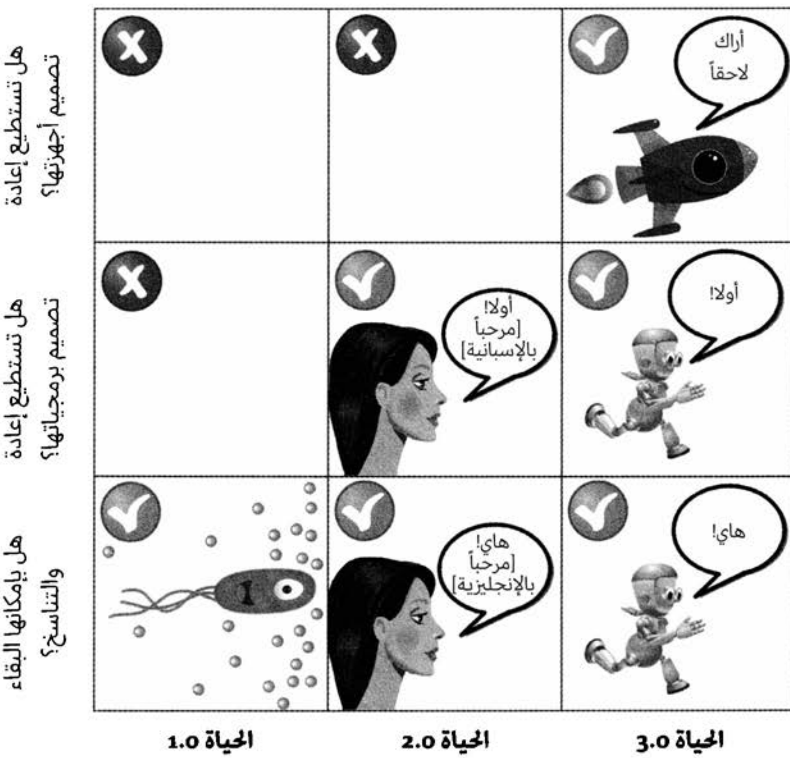
مثل كوننا نفسه، تدريجيّاً صارت الحياة أكثر تعقيداً وإثارة للاهتمام،* وكما سأوضح الآن، أعتقد أنه من المفيد تصنيف أشكال الحياة إلى ثلاثة مستويات من التطور: الحياة 1.0، و2.0 و 3.0، وقد لخّصت هذه المستويات الثلاثة في الشكل 1.1.

ما زال السؤال حول كيف ومتى وأين ظهرت الحياة لأول مرة في عالمنا سؤالاً مطروحاً. ولكن هناك أدلة قوية على أن الحياة هنا على الأرض ظهرت منذ نحو أربعة بلايين سنة. وسرعان ما صار كوكبنا يعج بمجموعة متنوعة من أشكال الحياة. وقد استطاع أنجحها الذي سرعان ما تفوق على البقية، أن يستجيب لبيئته بطريقة ما. وخصوصاً، كانت أفراده على شاكلة ما يسميه علماء الحاسوب «عوامل ذكية»: كيانات تجمع المعلومات حول بيئتها من أجهزة الاستشعار

* لماذا تزداد الحياة تعقيداً؟ التطور يكافئ الحياة المعقدة بدرجة كافية للتنبؤ واستغلال الأمور المنتظمة في بيئتها، لذلك في بيئة أكثر تعقيداً، ستتطور حياة أكثر تعقيداً وذكاء. بعد ذلك ستولد هذه الحياة الأكثر ذكاء بيئةً أكثر تعقيداً لأشكال الحياة المتنافسة التي بدورها تتطور لتكون أكثر تعقيداً؛ مما يؤدي في النهاية إلى نشوء نظام إيكولوجي لحياة معقدة جدّاً.

ثم تُعالج هذه المعلومات لتحديد كيفية التصرف مرة أخرى في بيئتها. وقد يتضمن ذلك معالجة معلومات معقدة جدّاً، مثل استخدام معلومات من عينيك وأذنيك لتقرر ماذا تقول في المحادثة، ولكن يمكن أن يشمل أيضاً الأجهزة Hardware والبرمجيات Software البسيطة جدّاً.

مثلاً، العديد من البكتيريا لديها جهاز استشعار لقياس تركيز السكر في السائل من حولها، وقد تسبح باستخدام بُنى على شكل آلة دفع تُسمى السوط Flagella. والمُعدات التي تربط أجهزة الاستشعار بالسوط قد تنفذ الخوارزمية البسيطة ولكن المفيدة التالية: «إذا رصد مستشعر تركيز السكر قيمة أقل مما كان عليه قبل بضع ثوان، إذاً أعكس اتجاه دوران السوط بحيث يمكنني تغيير الاتجاه».



الشكل 1.1: مراحل الحياة الثلاث: التطور البيولوجي، والتطور الثقافي، والتطور التكنولوجي. يتعذر على الحياة 1.0 إعادة تصميم أي من أجهزتها أو برمجياتها خلال حياتها: كلتاهما تتحدد من قبل الحمض النووي DNA الخاص بها. وتتغير فقط من خلال التطور على مدى أجيال عديدة، وعلى النقيض من ذلك، تستطيع الحياة 2.0 إعادة تصميم جزء كبير من برمجياتها: يستطيع البشر تعلم مهارات جديدة معقدة - مثلاً، اللغات والألعاب الرياضية والمهن - ويمكنهم تحديث رؤيتهم وأهدافهم بشكل أساسي. أما الحياة 3.0، التي لا توجد على وجه الأرض حتى الآن؛ فيمكنها إعادة تصميم برمجياتها وأيضاً أجهزتها، بدلاً من الاضطرار إلى الانتظار حتى تتطور تدريجيّاً عبر أجيال.

لقد تعلمت -أنت- كيفية التحدث ومهارات أخرى لا حصر لها. ولكن من ناحية أخرى، ليست البكتيريا من المتعلمين الرائعين، فحمضها النووي DNA يُحدِّد ليس فقط تصميم الأجهزة الخاصة بها، مثل أجهزة استشعار السكر والسوط، ولكن أيضاً تصميم برمجياتها. فهي لا تتعلم أبداً السباحة نحو السكر، بدلاً من ذلك، تُرمَّز الخوارزمية في حمضها النووي DNA من البداية. كانت هناك بالطبع عملية تعلّم من نوع ما، ولكنها لم تحدث خلال فترة حياة تلك البكتيريا المعينة. وبدلاً من ذلك، حدث ذلك خلال التطور السابق لأنواع البكتيريا، من خلال عملية التجربة والخطأ البطيئة التي امتدت على مدار عدة أجيال. ففضَّل الانتخاب (الانتقاء) الطبيعي Natural Selection تلك الطفرات Mutations العشوائية من الحمض النووي DNA التي حسَّنت استهلاك السكر، بعض من هذه الطفرات ساعد على تحسين تصميم السوط والأجهزة الأخرى، في حين حسَّنت الطفرات الأخرى نظام معالجة المعلومات البكتيرية التي تطبق خوارزمية العثور على السكر وغيرها من البرمجيات.

مثل هذه البكتيريا هي مثال لما أسميه "الحياة 1.0" Life 1.0: **الحياة التي تتطور فيها كل من الأجهزة والبرمجيات بدلاً من أن تكون مُصمَّمة**. ومن ناحية أخرى، أنت وأنا من أمثلة "الحياة 2.0": **الحياة التي تتطور معداتها، ولكن برمجياتها مُصمَّمة إلى حد كبير**، وما أعنيه بـ «برمجياتك» هو جميع الخوارزميات والمعرفة التي تستخدمها لمعالجة المعلومات الواردة من حواسك وتقرر ما يجب فعله - كل شيء من القدرة على التعرف على أصدقائك عندما تراهم إلى قدرتك على المشي والقراءة والكتابة والحساب والغناء وسرد النكات.

لم تكن قادراً على تنفيذ هذه المهام عند ولادتك، لذلك فجميع هذه البرمجيات تُبرمَج في عقلك بوقت لاحق من خلال العملية التي نطلق عليها اسم التعلُّم. وبينما يكون منهج الطفولة الخاص بك مصمماً بشكل كبير من قبل عائلتك ومعلميك، الذين يقررون ما يجب أن تتعلمه، فإنك تكتسب تدريجيّاً المزيد من القوة لتصميم برمجياتك الخاصة بك. وربما تسمح لك مدرستك باختيار لغة أجنبية: هل ترغب في تثبيت وحدة برمجية في عقلك تمكنك من التحدث بالفرنسية أو أخرى تمكنك من التحدث باللغة الإسبانية؟ هل تريد أن تتعلم لعب التنس أو الشطرنج؟ هل ترغب في الدراسة لتصبح شيف طبخ أو محامياً أو صيدلانياً؟ هل تريد معرفة المزيد عن الذكاء الاصطناعي ومستقبل الحياة من خلال قراءة كتاب عنه؟

قدرة الحياة 2.0 هذه على تصميم برمجياتها تجعلها أكثر ذكاء من الحياة 1.0. والذكاء العالي يتطلب الكثير من الأجهزة (المصنوعة من الذرات) والكثير من البرمجيات (مصنوعة من البتات). وحقيقة أن معظم أجهزتنا البشرية تضاف بعد الولادة (من خلال النمو) هي أمر مفيد، نظراً لأن حجمنا النهائي لا يقتصر على عرض قناة ولادة أُمِّنا. بالمثل، فإن حقيقة أن معظم برمجياتنا البشرية تضاف بعد الولادة (من خلال التعلم) مفيدة، نظراً لأن الذكاء النهائي لدينا لا يقتصر على مقدار المعلومات التي قد

تنتقل إلينا عند الحمل عبر الحمض النووي DNA، كما في الحياة 1.0. أَزِنُ نحو 20 ضعف ما ولدت به، والمشتبكات العصبية Synaptic connections التي تربط الخلايا العصبية (العصبونات) Neurons في دماغي تُخزِّن نحو مئة ألف ضعف المعلومات في الحمض النووي DNA الذي ولدت به، إذ تخزن مشتبكاتك العصبية جميع معارفك ومهاراتك بما يعادل 100 تيرابايت تقريباً من المعلومات، في حين أن المعلومات المُخزَّنة في حمضك النووي DNA هي نحو غيغابايت واحد فقط، وهو بالكاد يكفي لتخزين تنزيل فيلم واحد. لذلك من المستحيل جسديّاً أن يولد طفل يتحدث الإنجليزية بشكل كامل ومستعد لامتحانات القبول في كليته: لا يمكن تحميل المعلومات مسبقاً في دماغه، نظرا لأن وحدة المعلومات الرئيسية التي حصل عليها من والديه (الحمض النووي) لا تمتلك سعة التخزين الكافية.

والقدرة على تصميم برمجياتها لا تُمكِّن الحياة 2.0 من أن تكون أكثر ذكاء من الحياة 1.0 فحسب، بل أن تكون أكثر مرونة. فإذا تغيرت البيئة؛ يمكن للحياة 1.0 أن تتكيّف فقط عن طريق التطور البطيء لعدة أجيال. ولكن يمكن للحياة 2.0 أن تتكيف على الفور تقريباً، من خلال تحديث البرمجيات. مثلاً، البكتيريا التي تواجه المضادات الحيوية في كثير من الأحيان قد تُطور مقاومة الدواء على مدى أجيال عديدة، لكن البكتيريا الفردية لن تغير سلوكها على الإطلاق. وعلى النقيض من ذلك، فإن الفتاة التي تتعلم أنها مصابة بحساسية الفول السوداني ستغير سلوكها فوراً لتجنب تناول الفول السوداني. وهذه المرونة تمنح الحياة 2.0 ميزة أكبر على مستوى المجاميع السكانية: فعلى الرغم من أن المعلومات في الحمض النووي البشري لم تتطور بشكل كبير على مدار الخمسين ألف سنة الماضية، فإن المعلومات المُخزنة بشكل جماعي في أدمغتنا والكتب وأجهزة الحاسوب تفجّرت من حيث الحجم. ومن خلال تثبيت وحدة برمجية تمكننا من التواصل عبر لغة محكية متطورة، ضمنا أن المعلومات الأكثر فائدة المخزنة في دماغ شخص ما يمكن أن تُنسخ إلى أدمغة أخرى، والتي يحتمل أن تبقى على قيد الحياة حتى بعد موت الدماغ الأصلي. ومن خلال تثبيت وحدة برمجية تمكننا من القراءة والكتابة، أصبحنا قادرين على تخزين وتقاسم المزيد من المعلومات أكثر مما يستطيع الناس حفظه. ومن خلال تطوير برمجيات الدماغ القادرة على إنتاج التكنولوجيا (بمعنى، من خلال دراسة العلوم والهندسة)، مكَّنَّا الكثير من الأشخاص في العالَم من الوصول إليها بنقرات قليلة فقط.

لقد مكّنت هذه المرونة الحياة 2.0 من السيطرة على الأرض. فبعد أن تحرَّرت من قيودها الوراثية، استمرت المعرفة المشتركة للإنسانية في النمو بوتيرة متسارعة، إذ مكّن كل إنجاز خارق الإنجاز الخارق التالي: اللغة، فالكتابة، فالمطبعة، فالعلوم الحديثة، فأجهزة الحاسوب، فالإنترنت، وما إلى ذلك. وقد برزت البرمجيات باعتبارها القوة المهيمنة التي تُشكِّل مستقبلنا البشري؛ مما يجعل تطورنا البيولوجي البطيء جدّاً غير ذي أهمية تقريباً.

ولكن، على الرغم من أقوى التكنولوجيات المتوفرة لدينا اليوم، فإن جميع أشكال الحياة التي نعرفها تظل محدودة بشكل أساسي بفعل أجهزتها البيولوجية، فلا يمكن لأي

منها أن يعيش لمليون سنة، أو يحفظ الويكيبيديا كلها، أو يفهم كل العلوم المعروفة، أو يستمتع بالسفر في الفضاء من دون مركبة فضائية. ولا يمكن لأحد أن يحوِّل الكوزموس الذي لا حياة فيه إلى بيوسفير (عالم حيوي) Biosphere متنوع سيزدهر لبلايين أو تريليونات السنين، فيمكّن كوننا من أن يحقِّق إمكاناته ويستيقظ استيقاظاً تامّاً. وكل هذا يتطلب خضوع الحياة إلى عملية ترقية أخرى، لتصل إلى إصدار الحياة 3.0، الحياة التي لا يمكنها تصميم برمجياتها فحسب، بل أيضاً أجهزتها، بعبارة أخرى، الحياة 3.0 هي سيدة مصيرها، وخالية تماماً من قيوده التطورية.

غير أن الحدود بين مراحل الحياة الثلاث ضبابية بعض الشيء، إذا كانت البكتيريا هي الحياة 1.0 والبشر الحياة 2.0، عندئذ يمكنك تصنيف الفئران على أنها الحياة 1.1: إذ يمكنها تعلم الكثير من الأشياء، ولكن ليس بما يكفي لتطوير لغة أو اختراع الإنترنت. إضافة إلى ذلك، ولأنها تفتقر إلى اللغة، فإن ما تتعلمه يُفقد إلى حد كبير عندما تَنْفُق، ولا تنتقل هذه المعرفة إلى الجيل التالي. وبالمثل، يمكنك القول إن البشر اليوم يجب أن يُحتسبوا على أنهم الحياة 2.1: فيمكننا إجراء تحسينات بسيطة في الأجهزة مثل زرع الأسنان العضوية والركبتين وضبط النبضات، ولكن ليس شيئاً دراميّاً مثل إطالة الجسم بعشرة أضعاف أو الحصول على دماغ أضخم بألف مرة.

باختصار، يمكننا تقسيم تطور الحياة إلى ثلاث مراحل، تمتاز عن بعضها البعض بمدى مقدرة الحياة على تصميم نفسها:

- الحياة 1.0 (المرحلة البيولوجية): تُطوِّر أجهزتها Hardware وبرمجياتها Software
- الحياة 2.0 (المرحلة الثقافية): تُطوِّر أجهزتها وتُصمِّم الكثير من برمجياتها
- الحياة 3.0 (المرحلة التكنولوجية): تصمم أجهزتها وبرمجياتها

بعد 13.8 بليون سنة من التطور الكوزموسي، تسارعت وتيرة التطور بشكل كبير هنا على الأرض: لقد انبثقت الحياة 1.0 قبل أربعة بلايين سنة، ووصلنا إلى مرحلة الحياة 2.0 (نحن البشر) قبل نحو مئة ألف عام. ويعتقد العديد من الباحثين في مجال الذكاء الاصطناعي أن الحياة 3.0 قد تتحقق خلال القرن القادم، وربما حتى خلال حياتنا، وستولد بفعل التقدم في الذكاء الاصطناعي، ماذا سيحدث وما الذي سيعنيه هذا بالنسبة إلينا؟ هذا هو موضوع هذا الكتاب.

الجدال

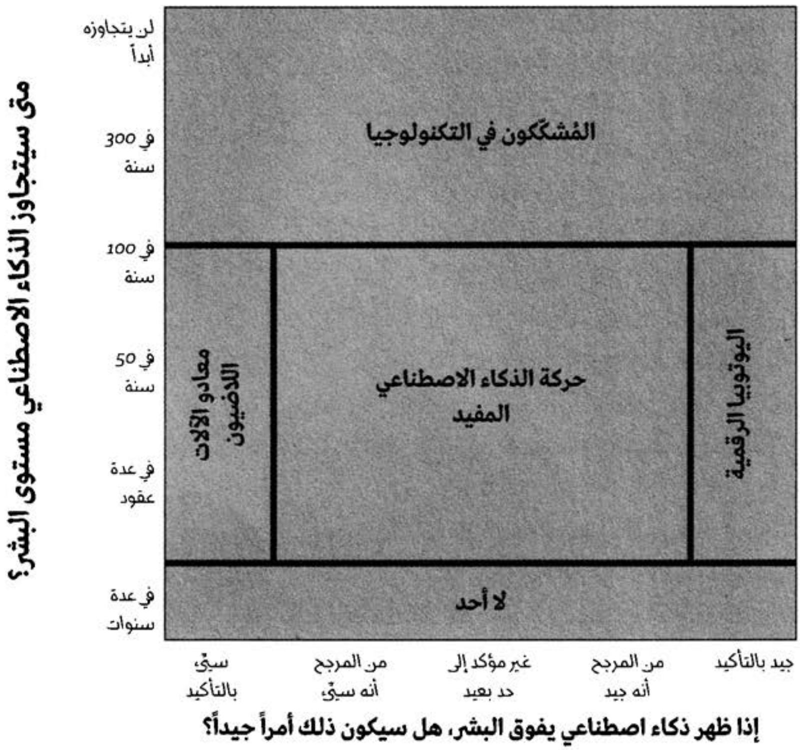
إن هذه المسألة، هي مسألة مثير لجدل رائع؛ إذ اختلف الباحثون الرائدون في مجال الذكاء الاصطناعي في العالم بحماس ليس فقط في توقعاتهم، ولكن أيضاً في ردود أفعالهم العاطفية، التي تتراوح بين التفاؤل الواثق والقلق الشديد. فهم غير مُجمعين

حتى بخصوص الأسئلة قصيرة المدى مثل الأثر الاقتصادي والقانوني والعسكري للذكاء الاصطناعي. بل تزداد اختلافاتهم عندما نوسع الأفق الزمني ونسأل عن **الذكاء الاصطناعي العام** ***Artificial general intelligence*** (اختصاراً: الذكاء AGI) - وخصوصا حول الذكاء AGI الذي يسعى إلى الوصول إلى المستوى البشري وما بعده؛ مما يمكّن من الوصول إلى الحياة 3.0. ويمكن **للذكاء الاصطناعي العام** تحقيق أي هدف تقريباً، بما في ذلك التعلم، على النقيض من الذكاء الضيق لبرنامج لعب الشطرنج.

ومن المثير للاهتمام أن الجدل حول الحياة 3.0 لا يدور حوله سؤال واحد لكن سؤالان منفصلان: متى وماذا؟ متى (إذا حدث) سيحدث، وماذا سيعني للبشرية؟ وفي رأيي، هناك ثلاث مدارس فكرية متميزة يجب أن تؤخذ على محمل الجد، لأن كلاً منها يضم عدداً من الخبراء الرائدين في العالم، وكما هو موضح في الشكل 2.1، فإنني أصنفهم **كطوباويين رقميين** ***Digital utopians***، **ومتشككين في التكنولوجيا** ***Technoskeptics***، **وأعضاء حركة الذكاء الاصطناعي المُفيد** ***Members of the beneficial-AI movement***، على التوالي. واسمحوا لي بأن أقدم إليكم بعضاً من أبطالها البليغين.

طوباويون رقميون

عندما كنت طفلاً، تخيلت أن البليونيرات يختالون بالتباهي والغطرسة. وعندما التقيت للمرة الأولى بلاري بيج Larry Page في غوغل Google في عام 2008، حطم هذه الصور النمطية تماماً. فقد كان يرتدي ملابس غير رسمية: جينزاً وقميصاً عاديّاً جدّاً، ولا يتميز عن الآخرين في نزهة من نزهات المعهد MIT. وحديثه اللطيف والعميق وابتسامته الودودة وأنا أتحدث إليه جعلاني أشعر بالراحة لا الرهبة. وفي 18 يوليو 2015 التقينا معاً في حفلة في نابا فالي أقامها إيلون ماسك Elon Musk وزوجته، تالولا، وتحدثنا عن اهتمامات أطفالنا بالقاذورات. وأوصيت بالكتاب الأدبي الكلاسيكي **اليوم الذي جُنّ فيه جنون مؤخرتي** ***The Day My Butt Went Psycho***، بقلم أندي ريفيثس Andy Griffiths، وطلبه لاري على الفور. وجاهدتُ لأذكّر نفسي أن لاري قد يدخل التاريخ باعتباره الإنسان الأكثر تأثيراً على الإطلاق: فتخميني أنه إذا استولت الحياة الرقمية الفائقة على كوننا في حياتي؛ فسيكون ذلك بسبب قرارات لاري.



الشكل 2.1: معظم الجدل حول ذكاء اصطناعي قوي (يمكنه أن يضاهي الإنسان في أي مهمة معرفية) تركز حول سؤالين: متى سيحدث ذلك؛ (إذا كان سيحدث)، وهل سيكون شيئاً جيداً للبشرية؟ يتفق المشككون في التقنية والطوباويون الرقميون على أنه لا ينبغي لنا أن نقلق، ولكن لأسباب مختلفة جداً: فالفريق الأول مقتنع بأن الذكاء الاصطناعي على المستوى البشري Human- Artificial general intelligence (اختصاراً: الذكاء AGI) لن يتحقق في المستقبل المنظور، في حين يعتقد الفريق الأخير أنه سيتحقق بل ومن المضمون أنه سيكون أمراً جيداً. وتشعر الحركة الذكاء الاصطناعي المُفيد بأن القلق مُبرَّر ومفيد؛ لأن أبحاث سلامة الذكاء الاصطناعي والمناقشات الدائرة حوله تزيد الآن من فرص التوصل إلى نتيجة جيدة. أما أتباع الحركة اللاضية (معادو الآلات) Luddite؛ فهم مقتنعون بأن النتيجة ستكون سيئة ويعارضون الذكاء الاصطناعي. وهذا الشكل البياني مستوحى جزئيّاً من تيم أوربان.[1]

وانتهى الحوار إلى تناول العشاء مع زوجتينا لوسي و ميا ومناقشة ما إذا كانت الآلات بالضرورة واعية، وهي القضية التي جادل لاري في أنها موضوع يصرف الانتباه عن القضية المهمة. وفي وقت لاحق من تلك الليلة، وبعد تناول المشروبات، دار نقاش طويل ومحتدم بينه وبين إيلون حول مستقبل الذكاء الاصطناعي وما يجب القيام به. ومع دخولنا في الساعات الأولى من صباح اليوم التالي، استمرت دائرة المشاهدين والمتفرجين بالنمو، وقدم لاري دفاعاً حارّاً عن الموقف الذي أُحبُّ-أنا- التفكير فيه على أنه اليوتوبيا الرقمية Digital utopianism: إن الحياة الرقمية هي الخطوة التالية

الطبيعية والمرغوب فيها بالتطور الكوزموسي، وإننا إذا منحنا العقول الرقمية الحرية بدلاً من محاولة إيقافها أو استعبادها، فمن المؤكد تقريباً أن تكون النتيجة جيدة. وأعتقد أن لاري أكثر دعاة اليوتوبيا الرقمية تأثيراً، وقد جادل في أنه إذا كانت الحياة ستنتشر في جميع أنحاء مجرتنا وما وراءها، وهو ما يعتقد أنه ينبغي أن يحدث، فينبغي أن يحدث ذلك في شكل رقمي. أما مخاوفه الرئيسية؛ فهي أن تسود الفوبيا الرقمية فتؤخر اليوتوبيا الرقمية و/أو تتسبب في أن يستولي الجيش على الذكاء الاصطناعي؛ فتكون في موقع يتعارض من شعار غوغل «لا تكن شريراً». وظل إيلون يدافع ويطلب إلى لاري أن يوضِّح حججه، مثل لماذا هو واثق من أن الحياة الرقمية لن تدمر كل شيء يهمنا. وفي بعض الأحيان اتهم لاري إيلون بأنه متحيز ضد النوع Specieist: ينظر إلى بعض أشكال الحياة على أنها أدنى فقط؛ لأنها تعتمد على السيليكون بدلاً من الكربون. وسنعود لاستكشاف هذه القضايا والحجج المثيرة للاهتمام بالتفصيل لاحقاً، وذلك بدءاً من الفصل الرابع.

على الرغم من أنه بدا أن لاري هزم بالكثرة في تلك الأمسية الصيفية الدافئة عند حمام السباحة، فإن اليوتوبيا الرقمية التي دافع عنها ببلاغة لها العديد من المؤيدين البارزين. فقد ألهم هانز مورافيك Hans Moravec، الداعي إلى الروبوتات والمبشر بالمستقبل، جيلاً كاملاً من الطوباويين الرقميين وذلك في كتابه الكلاسيكي **أطفال العقل *Mind Children*** الصادر عام 1988، وهو تقليد مستمر ومُحسَّن من قبل المخترع راي كورزويل Ray Kurzweil. كما قدم ريتشارد سوتون Richard Sutton، أحد رواد المجموعة الفرعية للذكاء الاصطناعي المعروفة بالتعلُّم التعزيزي Reinforcement learning، دفاعاً شغوفاً عن اليوتوبيا الرقمية في مؤتمرنا في بورتوريكو، وسأخبركم عن ذلك قريباً.

المشككون في التكنولوجيا

هناك مجموعة أخرى من المفكرين البارزين لا يساورهم القلق بشأن الذكاء الاصطناعي، ولكن لسبب مختلف تماماً: فهم يعتقدون أن تطوير ذكاء AGI يفوق البشر هو أمر صعب جدّاً، بحيث لن يحدث لمئات السنين، ومن ثم يرون القلق بشأنه الآن أمراً سخيفاً، أعتقد أن هذا هو الموقف المُشكِّك في التكنولوجيا، الذي عبّر عنه أندرو نغ Andrew Ng ببلاغة:

«إن الخوف من ارتفاع عدد الروبوتات القاتلة يشبه القلق بشأن الاكتظاظ السكاني على سطح المريخ»، وأندرو هو كبير العلماء في بايدو Baidu، شركة غوغل الصينية، وقد كرر هذه الحجة مؤخراً عندما تحدثت إليه في مؤتمر في بوسطن. كما أخبرني أنه يشعر بأن القلق بشأن خطر الذكاء الاصطناعي هو تشتيت ضار محتمل قد يبطئ تقدم الذكاء الاصطناعي. وقد عبَّر عن مشاعر مماثلة عدد من المُشكِّكين بالتكنولوجيا

الآخرين مثل رودني بروكس Rodney Brooks، الأستاذ السابق في المعهد MIT الذي طوّر المكنسة الكهربائية الروبوتية Roomba وروبوت باكستر Baxter الصناعي. وأجد من المثير للاهتمام أنه على الرغم من أن اليوتوبيا الرقمية والمشككين في التكنولوجيا يتفقون على أنه لا ينبغي لنا أن نقلق بشأن الذكاء الاصطناعي، فإنهم يتفقون على أمور أخرى قليلة. ويعتقد معظم اليوتوبيين أن مستوى الذكاء AGI قد يصل إلى المستوى البشري خلال العشرين أو المئة سنة المقبلة، وهو ما يطرحه المُشكِّكون في التكنولوجيا جانباً كحلمٍ هائمٍ. وهم غالباً ما يسخرون من التفرد المتنبِئ بأنه سيكون «بهجة الأذكياء المهووسين Geeks».* وفي ديسمبر 2014 قابلت رودني بروكس في حفل عيد ميلاد، وأخبرني أنه متأكد بنسبة 100% من أنه لن يحدث في حياتي، «هل أنت متأكد من أنك لا تعني 99%؟»، سألت في رسالة إلكترونية لمتابعة النقاش، وأجاب، «لا 99% هزيلة، 100%، لن يحدث ذلك».

حركة الذكاء الاصطناعي المفيد

عندما التقيت ستيوارت راسل Stuart Russell لأول مرة في أحد المقاهي بباريس في يونيو 2014، صدمني كجنتلمان بريطاني نمطي. فقد كان بليغاً وعميقاً ولطيف الحديث. ولكن مع بريق المغامرة في عينيه، بدا لي تجسيداً معاصراً لفيلياس فوغ، بطل طفولتي من رواية جول فيرن Jules Verne الكلاسيكية الصادرة عام 1873: **حول العالم في 80 يوماً** ***Around the World in 80 Days***. وعلى الرغم من أنه كان أحد أشهر الباحثين في مجال الذكاء الاصطناعي، فإنه شارك في تأليف الكتاب المدرسي القياسي حول هذا الموضوع. وسرعان ما جعلني تواضعه ودفئه يشعرني بالراحة، وشرح لي كيف أن التقدم في الذكاء الاصطناعي أقنعه بأن الوصول إلى ذكاء AGI بمستوى بشري خلال هذا القرن كان احتمالاً حقيقيّاً. وعلى الرغم من أنه كان يأمل بذلك، فإنه لا يمكن ضمان التوصل إلى نتيجة جيدة، فنحن بحاجة إلى الإجابة عن أسئلة حاسمة أولاً، وهي أمر صعب جدّاً، ويجب أن نبدأ البحث عنها الآن، بحيث تكون لدينا الإجابات جاهزة بحلول الوقت الذي نحتاج إليها فيه.

اليوم، تعتبر آراء راسل سائدة إلى حد ما، وتنكبُّ مجموعات كثيرة حول العالم على أبحاث السلامة في مجال الذكاء الاصطناعي من النوع الذي يدعو إليها. ولكن هذه لم تكن الحال دائماً، فقد أشارت مقالة نشرت في صحيفة **ذا واشنطن بوست** ***The Washington Post*** إلى أن عام 2015 هو السنة التي صارت فيها أبحاث سلامة الذكاء الاصطناعي هي الأبحاث السائدة. وقبل ذلك، غالباً ما أساء الباحثون الرئيسيون

* يستخدم المؤلف عبارة «Rapture of the geeks»، وهو عنوان كتاب من تأليف أستاذ القانون ريتشارد دولينغ، يتفوق فيه الذكاء الاصطناعي على البشري. [المترجم]

في الذكاء الاصطناعي فهم أي حديث عن مخاطر الذكاء الاصطناعي، وطرحوه جانبا كتخويف ضد الآلة يهدف إلى عرقلة تقدم الذكاء الاصطناعي. وكما سنستكشف في الفصل الخامس، كانت المخاوف الشبيهة بمخاوف راسل قد ظهرت لأول مرة منذ أكثر من نصف قرن من قبل رائد الحاسوب آلان تورينغ Alan Turing وعالم الرياضيات إيرفينغ جيه. غوود Irving J. Good الذي عمل مع تورينغ لكسر شيفرة الرموز الألمانية خلال الحرب العالمية الثانية. وفي العقد الماضي كان البحث في هذه المواضيع تجربه في الغالب حفنة من المفكرين المستقلين الذين لم يكونوا باحثين محترفين في مجال الذكاء الاصطناعي، مثلاً اليعازر يودكوفسكي Eliezer Yudkowsky، ومايكل فاسار Michael Vassar، ونيك بوستروم Nick Bostrom. وقد كان لدراساتهم تأثير ضئيل في أعمال معظم الباحثين الرئيسيين في الذكاء الاصطناعي، والذين كانوا يميلون إلى التركيز على مهامهم اليومية في جعل أنظمة الذكاء الاصطناعي أكثر ذكاء بدلاً من التفكير في عواقب النجاح على المدى البعيد. ولكن من بين الباحثين في الذكاء الاصطناعي، كنت أعرف عدداً ممن يساورهم بعض القلق، وقد تَردَّد كثيرون في التعبير عن ذلك خوفاً من أن ينظر إليهم كمثيرين للفزع معادين للتكنولوجيا.

وشعرت بأن هذا الوضع المستقطب بحاجة إلى التغيير، بحيث يمكن لمجتمع الذكاء الاصطناعي بكامله أن ينضم إلى الحوار حول كيفية تطوير الذكاء الاصطناعي. ولحسن الحظ، لم أكن وحدي، ففي ربيع عام 2014، أسست منظمة غير ربحية تدعى معهد مستقبل الحياة Future of Life Institute (http://futureoflife.org -اختصاراً المعهد FLI- مع زوجتي ميا، وصديقي عالم الفيزياء أنتوني أُغيري Anthony Aguirre، وطالبة الدراسات العليا في هارفارد فيكتوريا كراكوفنا Victoria Krakovna، ومؤسس سكايب يان تالين Jaan Tallinn. كان هدفنا بسيطاً: المساعدة على ضمان وجود مستقبل للحياة وأن يكون رائعاً قدر الإمكان. وعلى الخصوص، شعرنا بأن التكنولوجيا تمنح الحياة القوة، إما لتزدهر كما لم يسبق لها ذلك أو أن تدمر نفسها بنفسها، وقد فضلنا الرؤية الأولى.

كان اجتماعنا الأول عبارة عن جلسة عصف ذهني Brainstorming في منزلنا في 15 مارس 2014، مع نحو 30 طالباً وأستاذاً ومفكراً آخرين من منطقة بوسطن. وكان هناك إجماع واسع على أنه على الرغم من أننا يجب أن نولي اهتماماً بالتكنولوجيا الحيوية، والأسلحة النووية وتغير المناخ، فإن الهدف الرئيسي الأول لنا هو المساعدة على جعل الأبحاث حول سلامة الذكاء الاصطناعي أبحاثاً سائدة. واقترح زميلي الفيزيائي من المعهد MIT، فرانك ويلكزك Frank Wilczek، الحاصل على جائزة نوبل لمساعدته على معرفة كيفية عمل الكواركات، أن نبدأ بكتابة مقالة افتتاحية للفت الانتباه إلى القضية وجعل تجاهلها أكثر صعوبة. فتواصلت مع ستيوارت راسل (الذي لم أكن قد قابلته بعدُ) وزميلي في الفيزياء ستيفن هوكينغ Stephen Hawking، اللذين اتفقا على الانضمام إليّ وإلى فرانك كمؤلفين مشاركين.

وبعد عدد من التعديلات، رفضت صحيفة **ذَ نيويورك تايمز** ***The New York Times*** مقالتنا المشتركة وكذلك العديد من الصحف الأمريكية الأخرى، لذلك نشرناها على مدونتي في **هافينغتون بوست** ***Huffington Post***. ومما سرَّني، أن أريانا هوفينغتون نفسها أرسلت رسالة إلكترونية وقالت: «سعداء أن نتلقى المقالة! *سننشرها على الصفحة #1*!» وضع المقالة في الجزء العلوي من الصفحة الأولى أثار موجة من التغطية الإعلامية تتناول سلامة الذكاء الاصطناعي التي استمرت لبقية العام. وردد صداها إيلون ماسك وبيل غيتس Bill Gates وقادة التكنولوجيا الآخرين. وفي ذلك الخريف صدر كتاب نيك بوستروم بعنوان **الذكاء الفائق** ***Superintelligence*** وأذكى النقاش العام المتنامي.

وكان الهدف التالي من حملتنا للذكاء الاصطناعي المفيد -ضمن أعمال المعهد FLI- هو جذب الباحثين الرائدين في العالم من الذكاء الاصطناعي إلى مؤتمر يمكن فيه إزالة سوء التفاهم، وإيجاد توافق في الآراء، ووضع خطط بناءة. كنا نعلم أنه سيكون من الصعب إقناع مثل هذا الحشد اللامع بالقدوم إلى مؤتمر ينظمه غرباء لا يعرفونهم، خاصة إذا أخذنا بالاعتبار الموضوعَ المثير للجدل. لذلك حاولنا بأقصى ما نستطيع: فحظرنا وسائل الإعلام من الحضور، وعقدنا المؤتمر في منتجع شاطئي في يناير (في بورتوريكو)، وجعلنا التسجيل مجانيّاً (بفضل سخاء يان تالين، وأخترنا أكثر عنوان غير مثير للفزع يمكننا التوصل إليه: «مستقبل الذكاء الاصطناعي: الفرص والتحديات» The Future of AI: Opportunities and Challenges. والأهم من ذلك، أننا تعاوننا مع ستيوارت راسل، الذي بفضله تمكنا من تشكيل اللجنة المنظمة لتشمل مجموعة من قادة الذكاء الاصطناعي من الأوساط الأكاديمية والصناعية، بمن في ذلك ديميس هاسابيس Demis Hassabis من ديب مايند DeepMind من غوغل، الذي أظهر أن الذكاء الاصطناعي قادر على أن يهزم البشر حتى في لعبة غو Go. وكلما تعرفت على ديميس أكثر، أدركت أنه كان لديه طموح ليس فقط لجعل الذكاء الاصطناعي قويّاً، ولكن أيضاً لجعله مفيداً.

وكانت النتيجة اجتماعاً رائعاً للعقول (الشكل 3.1)، انضم إلى الباحثين في الذكاء الاصطناعي كبار الاقتصاديين، وعلماء القانون، وقادة التكنولوجيا (بمن في ذلك إيلون ماسك) والمفكرين الآخرين (بمن في ذلك فيرنور فينغ Vernor Vinge، الذي صاغ مصطلح «حدث منفرد» Singularity، وهو محور الفصل الرابع). وتجاوزت النتائج حتى أكثر توقعاتنا تفاؤلاً: وعلى الرغم من هذا الموضوع المثير للجدل، ظهر إجماع رائع لُخِّص في رسالة مفتوحة[2] انتهى بها الأمر إلى توقيع أكثر من ثمانية آلاف شخص بمن في ذلك العاملون في الذكاء الاصطناعي عليها. وكان جوهر الرسالة هو أنه ينبغي إعادة تعريف هدف الذكاء الاصطناعي: يجب ألا يكون الهدف هو خلق ذكاء غير موجه، ولكن ذكاء مفيد. وذكرت الرسالة أيضاً قائمة مفصلة بمواضيع البحث التي وافق المشاركون في المؤتمر على تعزيزها لهذا الهدف. وهكذا صارت حركة الذكاء الاصطناعي المفيد هي الاتجاه السائد. وسنتابع التقدم لاحقاً في الكتاب.



الشكل 3.1: جَمَع مؤتمرُ بورتوريكو في يناير 2015 مجموعة مُتميِّزة من الباحثين في مجال الذكاء الاصطناعي والمجالات ذات الصلة. الصف الخلفي، من اليسار إلى اليمين: توم ميتشيل، شيين أو، هيغيارتيغ برايس، شاميل شانداريا، يان تالين، ستيوارت راسل، بيل هيبارد، بليز أوغيورا إي آركاس، أندريه ساندبيرغ، دانييل ديوي، ستيورات آرمسترونغ، لوك مولهاوسر، توم ديتريتش، مايكل أوزبورن، جيمس منيكا، أجاي أغراوال، ريتشارد ملاح، نانسي تشانغ، ماثيو بوتمان، آخرون. ومن اليسار إلى اليمين: مارلين تومسون، ريتش ساتون، أليكس ويسنر-غروس، سام تيلر، توبي أُورد، جوشوا باخ، كاتيا غريس، أدريان ويلر، هيثر روف-بيركينز، ديليب جورج، شين ليغ، دميس هاسابيس، ويندل والاش، كريستينا تشوي، إيليا سوتسكيفر، كنت ووكر، سيسيليا تيلي، نيك بوستروم، إيريك برينجولفسون، ستيف كروسان، مصطفى سليمان، سكوت فينيكس، نيل جاكوبستين، موري شانهان، روبين هانسون، فرانشيسكا روسي، نيت سورس، إيلون ماسك، أندرو ماكْفي، بارت سيلمان، مايكل رايلي، آرون فان ديفندر، ماكس تيغمارك، مارغرت بودن، جوشوا غرين، بول كريستيانو، إيليزر يودكووسكي، ديفيد باركس، لوران أُورسو، جيه، بي، ستروبول، جيمس مور، شين ليغاسيك، ماسون هارتمان، هووي ليمبيل، ديفيد فالديك، جاكوب ستينهارت، مايكل فاسار، رايان كالو، سوزان يونغ، أويان إيفانز، ريفا-مليسا تيز، يانوس كريمار، جيف آندريه، فيرنون فينغ، أنثوني آوغيير. الجلوس: سام هاريس، توماس بوغيو، مارين سلوياسيك، فيكتوريا كراكوفنا، ميا شيتا-تيغمارك. خلف الكاميرا: أنثوني أوغيير (أيضاً عولجت الصورة بالفوتوشوب بالذكاء الإنساني للشخص الجالس إلى جواره).

أما الدرس الآخر المهم من المؤتمر؛ فهو: الأسئلة التي أثارها نجاح الذكاء الاصطناعي ليست مجدية فكريا فقط، بل إنها جوهرية أخلاقيّاً، لأن خياراتنا قد تؤثر في مستقبل الحياة بكامله. فقد كانت الدلالة الأخلاقية لخيارات البشرية الماضية كبيرة في بعض الأحيان، ولكنها كانت محدودة دائماً: لقد تعافينا حتى من أسوأ الأوبئة، وحتى أكبر الإمبراطوريات تنهار في نهاية المطاف. وقد أدركت الأجيال السابقة أن الشمس بكل تأكيد ستشرق غداً، وكذلك البشر في الغد وهم يعالجون آفات دائمة مثل الفقر والمرض والحرب. ولكن بعض المتحدثين في بورتوريكو جادلوا في أنه هذه المرة قد يكون الأمر مختلفا: إذ يقولون إننا لأول مرة قد نطور تكنولوجيا قوية بما يكفي لوضع حد نهائي لهذه الآفات - أو لإنهاء الإنسانية نفسها. وقد نخلق مجتمعات مزدهرة لم يسبق لها مثيل، سواء على الأرض أو ربما نصل إلى ما وراء الأرض، أو نُقيم دولة ترصُّد عالمَية -مثل أعمال كافكا- قوية جداً بحيث لا يمكن إسقاطها أبداً.



الشكل 4.1: على الرغم من أن وسائل الإعلام غالباً ما تصور إيلون ماسك على أنه على خلاف مع مجتمع الذكاء الاصطناعي، فإن هناك في الحقيقة توافقاً واسعاً في الآراء حول ضرورة إجراء أبحاث تتعلق بسلامة الذكاء الاصطناعي. التقطت هذه الصورة في 4 يناير 2015، ويظهر فيها توم ديتريتش- رئيس جمعية تقدم الذكاء الاصطناعي- وهو يشاطر إيلون سروره بالبرنامج الجديد لأبحاث سلامة الذكاء الاصطناعي الذي تعهد إيلون بتمويله قبل لحظات [من التقاط الصورة]، وخلفهما تقف مُؤَسِّستا المعهد FLI ميا شيتا-تيغمارك وفيكتوريا كراكوفنا.

مفاهيم خاطئة

عندما غادرت بورتوريكو، كنت مقتنعاً بضرورة استمرار الحوار الذي بدأناه حول مستقبل الذكاء الاصطناعي؛ لأنه أهم حوار في عصرنا.* إنه حوار حول مستقبلنا الجمعي جميعاً، لذا لا ينبغي أن يقتصر على الباحثين في الذكاء الاصطناعي فحسب. ولهذا السبب كتبتُ هذا الكتاب: كتبته على أمل أن تنضم أنت، يا عزيزي القارئ، إلى هذا الحوار، فأي نوع من المستقبل تريد؟ هل يجب علينا تطوير أسلحة قاتلة مستقلة استقلالاً تاماً؟ ماذا تريد أن يحدث مع أتمتة العمل؟ ما النصيحة المهنية التي ستقدمها لأطفال اليوم؟ هل تفضل نشوء وظائف جديدة لتحل محل الوظائف القديمة، أو مجتمع من العاطلين عن العمل يتمتع فيه الجميع بحياة الترفيه والثروة المُنتجة بالآلات؟ وفي المستقبل، هل تريد منّا تطوير الحياة 3.0 ونشرها عبر الكوزموس بأسره؟ هل سنسيطر على الآلات الذكية أم

* حوار الذكاء الاصطناعي مهم من حيث الإلحاح والتأثير. ومقارنةً بتغير المناخ الذي قد يحدث دماراً في خمسين إلى مئتي عام يتوقع العديد من الخبراء أن يكون للذكاء الاصطناعي تأثير أكبر خلال عقود - ومن المحتمل أن يوفر لنا تكنولوجيا للتخفيف من آثار تغير المناخ. ومقارنة بالحروب والإرهاب والبطالة والفقر والهجرة والعدالة الاجتماعية، سيكون لصعود الذكاء الاصطناعي تأثير عام أكبر - في الواقع، سنستكشف في هذا الكتاب الكيفية التي يمكنه بها السيطرة على ما يحدث في جميع هذه القضايا، للأفضل أو للأسوأ.

ستسيطر علينا؟ هل ستحل الآلات الذكية محلنا أو تتعايش معنا أو تندمج معنا؟ ماذا يعني أن تكون إنساناً في عصر الذكاء الاصطناعي؟ ماذا تريد أن يعني ذلك، وكيف يمكننا أن نجعل المستقبل أن يصبح على هذه الشاكلة؟

الهدف من هذا الكتاب هو مساعدتك على الانضمام إلى هذا الحوار. فكما ذكرت، هناك مناظرات رائعة يتجادل فيها كبار خبراء العالَم، لكنني رأيت أيضاً أمثلة كثيرة على الجدل الزائف الممل الذي يسيء الناسُ فيه فهمَ بعضهم البعض أو يتحدثون عن موضوعات مختلفة دون أن يدركوا ذلك. لنساعد أنفسنا على التركيز على الخلافات المثيرة للاهتمام والأسئلة المفتوحة، وليس على سوء الفهم، فلنبدأ بتوضيح بعض أكثر المفاهيم الخاطئة شيوعاً.

هناك العديد من التعاريف المتنافسة في الاستخدام الشائع لمصطلحات مثل "حياة" Life و "ذكاء" Intelligence و"وعي" Consciousness. وينشأ العديد من المفاهيم الخاطئة لأن الأفراد لا يدركون أنهم يستخدمون كلمة ما بطريقتين مختلفتين. وللتأكد من أننا -أنا وأنت- لا نقع في هذا الفخ، أعددت ورقة غش في الجدول 1.1 توضح الكيفية التي استخدم بها مصطلحات رئيسية في هذا الكتاب. وبعض هذه التعاريف لن تُقدَّم وتوضَّح بشكل صحيح إلا في فصول لاحقة. يرجى ملاحظة أنني لا أدعي أن تعاريفي أفضل من أي شخص آخر، أنا ببساطة أريد تجنب الارتباك من خلال توضيح ما أعنيه. سترى أني أسعى عموماً إلى تعاريف واسعة تتجنب التحيز المتمركز حول الإنسان، ويمكن تطبيقها على الآلات، وكذلك البشر. يرجى قراءة ورقة الغش الآن، والعودة إليها والتحقق منها لاحقاً إذا شعرت بالحيرة من كيفية استخدام إحدى كلماتها - خاصة في الفصول 4-8.

ورقة غش المصطلحات	
حياة Life	عملية يمكنها أن تحتفظ بتعقيدها والقدرة على تكرار نفسها
حياة Life 1.0	حياة تُطوِّر مُعداتها وبرمجياتها (المرحلة البيولوجية)
حياة Life 2.0	حياة تُطوِّر مُعدَّاتها ولكنها تصمم الكثير من برمجياتها (المرحلة الثقافية)
حياة Life 3.0	حياة تصمم مُعدَّاتها وبرمجياتها (المرحلة التكنولوجية)
ذكاء Intelligence	القدرة على تحقيق الأهداف المعقدة
ذكاء اصطناعي Artificial Intelligence (AI)	ذكاء غير بيولوجي المنشأ
ذكاء متخصص Narrow intelligence	القدرة على تحقيق مجموعة ضيقة من الأهداف، مثلاً، لعب الشطرنج أو قيادة السيارة
ذكاء عام General intelligence	القدرة على تحقيق أي هدف تقريباً، بما في ذلك التعلم

ذكاء عالمي Universal intelligence	القدرة على اكتساب الذكاء العام نظراً للقدرة على الوصول إلى البيانات والموارد
ذكاء اصطناعي عام [المستوى البشري] Humawn- level] Artificial] (General Intelligence (AGI	القدرة على إنجاز أي مهمة إدراكية على الأقل بالدرجة نفسها التي يحققها البشر
ذكاء اصطناعي على المستوى البشري Human- level AI	الذكاء AGI
ذكاء اصطناعي قوي	الذكاء AGI
ذكاء فائق	ذكاء عام يفوق مستوى البشر
حضارة Civilization	مجموعة متفاعلة من أشكال حياة ذكية
وعي Consciousness	تجربة شخصية
الكيفيات المحسوسة Qualia	الحالات الفردية للتجربة الشخصية
أخلاق Ethics	المبادئ التي تحكم الكيفية التي يجب أن نتصرف وفقها
علم الغائية Teleology	شرح الأشياء من حيث أهدافها أو أغراضها وليس أسبابها
السلوك المُوجَّه بالهدف Goal- oriented behavior	سلوك يمكن تفسيره من خلال تأثيره بسهولة أكثر أكثر من تفسيره من خلال سببه
امتلاك هدف Having a goal	عرض السلوك الموجّه للهدف
امتلاك غرض Having a purpose	خدمة أهداف الشخص أو كيان آخر
ذكاء اصطناعي ودود Friendly AI	ذكاء فائق أهدافه متوائمة مع أهدافنا
سايبورغ Cyborg	هجين إنسان-آلة
تفجر الذكاء Intelligence explosion	تحسين الذاتي متصل بسرعة مما يؤدي إلى ذكاء فائق
حدث منفرد Singularity	تفجر الذكاء
الكون Universe	منطقة الفضاء التي كان أمام الضوء وقت كافٍ للوصول إلينا خلال 13.8 بليون سنة منذ الانفجار الكبير Big bang

الجدول 1.1: العديد من حالات سوء الفهم حول الذكاء الاصطناعي سببها الأفراد الذين يستخدمون الكلمات أعلاه ليقصدوا بها أشياء مختلفة، إليكم ما أقصده في هذا الكتاب. (بعض هذه التعاريف لن تُقدّم وتوضَّح بشكل صحيح إلا في فصول لاحقة.)

إضافة إلى الارتباك في استخدام المصطلحات، شهدت أيضاً خروج العديد من الحوارات حول الذكاء عن مسارها بسبب مفاهيم خاطئة بسيطة، دعنا نوضّح أكثرها شيوعاً.

خرافات المسار الزمني

الخرافة الأولى هي بخصوص المسار الزمني في الشكل 2.1: كم من الوقت سيستغرق منا الأمر حتى تفوق الآلات بشكل كبير ذكاء المستوى البشري AGI على المستوى البشري؟ حول هذه النقطة، يشيع اعتقاد خاطئ بأننا نعرف الإجابة بدرجة كبيرة من اليقين.

إحدى الخرافات الشعبية هي أننا سنصل إلى ذكاء اصطناعي على المستوى البشري (الذكاء AGI) في هذا القرن. وفي الواقع، فإن التاريخ مفعم بفرط التفاؤل التكنولوجي، أين هي محطات الطاقة الاندماجية Fusion power plants والسيارات الطائرة التي وعدنا بأنها ستكون مستخدمة في وقتنا هذا؟ الذكاء الاصطناعي أيضاً كان موضوعاً لفرط التفاؤل التكنولوجي تكراراً ومراراً في الماضي، حتى من قبل بعض مؤسسي هذا الحقل: مثلاً، نجد أن جون مكارثي John McCarthy (الذي صاغ مصطلح «الذكاء الاصطناعي») ومارفن مينسكي Marvin Minsky وناثانيل روشستر Nathaniel Rochester وكلود شانون Claude Shannon أسّسوا لهذا التفاؤل المفرط في توقعات ما يمكن إنجازه خلال شهرين مع أجهزة الحاسوب من العصر الحجري حين كتبوا: «نقترح دراسة لمدة شهرين يُجريها عشرة أفراد على الذكاء الاصطناعي خلال صيف عام 1956 في كلية دارتموث Dartmouth College... وستجري محاولة لمعرفة كيفية جعل الآلات تستخدم اللغة، وتشكل الأفكار المُجرّدة والمفاهيم، وتحل أنواع المشكلات المقصورة الآن على البشر، وتحسن نفسها، ونعتقد أنه يمكن تحقيق تقدم كبير في واحدة أو أكثر من هذه المشكلات إذا عملت مجموعة مختارة من العلماء معاً بعناية خلال هذا الصيف».

وفي المقابل، نجد أيضاً الخرافة الشعبية القائلة إننا **لن** نصل إلى ذكاء اصطناعي على المستوى البشري (الذكاء AGI) في هذا القرن. وقد وضع الباحثون مجموعة واسعة من التقديرات حول إلى أي مدى نحن قريبون من تحقيق ذكاء اصطناعي يفوق المستوى البشري (الذكاء AGI). ولكننا بالتأكيد لا نستطيع أن نقول بثقة كبيرة إن احتمال تحقيق ذلك هذا القرن يعادل الصفر، نظراً للفشل الذريع لمثل هذه التنبؤات التقنية المُتشككة. مثلاً، قال إرنست رذرفورد Ernest Rutherford الذي يمكن القول إنه أعظم فيزيائي نووي في عصره، في عام 1933 وقبل أقل من 24 ساعة من اختراع ليو زيلارد Leo Szilard للتفاعل النووي المُتسلْسِل Nuclear chain reaction، إن الطاقة النووية كانت «كلاماً فارغاً». وفي عام 1956 وصف عالِمُ الفلك الملكي [في البلاط الملكي البريطاني] ريتشارد وولي Richard Woolley الحديث عن السفر إلى الفضاء بأنه «هراء محض». أما النسخة الأكثر تطرفاً من هذه الخرافة؛ فهي أن ذكاء اصطناعيّاً يفوق المستوى البشري (الذكاء AGI) لن يتحقق أبداً؛ لأنه مستحيل فعليا. ولكن، علماء الفيزياء يعلمون أن المخ يتكون من كواركات وإلكترونات مرتبة للعمل كحاسوب قوي، وأنه لا يوجد قانون فيزيائي يمنعنا من بناء مجموعات كوارك أكثر ذكاء.



الشكل 5.1: الخرافات الشائعة حول الذكاء الاصطناعي الفائق.

وقد أُجري العديد من الدراسات الاستقصائية التي سألت الباحثين في الذكاء الاصطناعي عن عدد السنوات التي يعتقدون -باحتمال يعادل 50% على الأقل- أننا نحتاج إليها لتحقيق ذكاء اصطناعي على المستوى البشري (الذكاء AGI)، وجاءت جميع هذه الاستطلاعات بالنتيجة نفسها: يختلف كبار خبراء العالم، لذلك فنحن ببساطة لا نعرف. مثلاً، عندما أجرى مثل هذا الاستطلاع على باحثي الذكاء الاصطناعي في مؤتمر الذكاء الاصطناعي في بورتوريكو، كان متوسط الإجابة بحلول عام 2055، ولكن بعض الباحثين خمنوا أننا نحتاج إلى مئات السنين أو أكثر.

وهناك أيضاً خرافة ذات صلة تفيد بأن الأشخاص الذين يقلقون من صعود الذكاء الاصطناعي يعتقدون أن تحقيق ذلك لن يستغرق سوى بضع سنوات. وفي الواقع، فإن معظم الأشخاص الذين يفصحون علانية عن قلقهم بشأن ذكاء اصطناعي يفوق المستوى البشري (الذكاء AGI) يعتقدون أن تحقيق ذلك لا يزال على بعد عقود من الزمن، لكنهم يقولون إنه ما دمنا غير متأكدين بنسبة 100% من أنه لن يحدث هذا القرن، فمن الحصافة أن نبدأ أبحاث السلامة الآن للإعداد لهذا الاحتمال. وكما سنرى في هذا الكتاب، فإن العديد من مشكلات السلامة صعبة جدّاً لدرجة أنها قد تستغرق عقوداً لحلها، لذا فمن الحكمة البدء بالبحث عنها الآن بدلاً من الالتفات إليها الليلة السابقة على اتخاذ بعض المبرمجين الذين يشربون مشروب الطاقة ريد بول Red Bull قرار إطلاق الذكاء الاصطناعي على المستوى البشري (الذكاء AGI).

خرافات الجدل

هناك اعتقاد خاطئ آخر شائع هو أن الأشخاص الوحيدين الذين لديهم مخاوف بشأن الذكاء الاصطناعي والذين يؤيدون البحث في مجال سلامة الذكاء الاصطناعي هم من أعداء الآلة الذين لا يعرفون الكثير عن الذكاء الاصطناعي. وعندما ذكر ستيوارت راسل ذلك خلال حديثه في بورتوريكو تَعالى ضحك الجمهور. وهناك اعتقاد خاطئ آخر مرتبط بهذا الاعتقاد الخاطئ يذهب إلى أن دعم أبحاث سلامة الذكاء هو موضع جدل كبير. وفي الواقع، من أجل دعم استثمار متواضع في أبحاث السلامة في مجال الذكاء الاصطناعي، لا يحتاج الأفراد إلى الاقتناع بأن المخاطر عالية، بل كل ما نحتاج إليه هو إقناعهم بأنها مخاطر لا يمكن تجاهل احتمالها، تماماً مثل الاستثمار البسيط في التأمين على المنازل يبرره احتمال لا يمكن تجاهله لاحتراق المنزل.

إن تحليلي الشخصي هو أن وسائل الإعلام جعلت النقاش حول سلامة الذكاء الاصطناعي يبدو أكثر إثارة للجدل مما هو عليه في الواقع. فالخوف يُروِّج للإقبال على موضوع ما، والمقالات التي تستخدم اقتباسات خارج السياق لإعلان الهلاك

الوشيك قد تولد المزيد من النقرات [على الموقع الإلكتروني] من تلك الاقتباسات الدقيقة والموزونة. ونتيجة لذلك، من المرجح أن يعتقد شخصان -لا يعرف أحدهما سوى موقف الآخر من اقتباسات وسائل الإعلام- أنهما يختلفان في وجهة النظر في حين أنهما قد يكونان متفقين. مثلاً، أحد المتشككين في التقنية الذين لا يعلمون إلا بموقف بيل غيتس من صحيفة بريطانية قد يظن خطأً أن بيل غيتس يعتقد أن الذكاء الفائق وشيك. وبالمثل، فإن شخصاً ما في حركة الذكاء الاصطناعي، لا يعرف شيئاً عن موقف أندرو نيغ Andrew Ng، باستثناء اقتباسه المذكور آنفاً حول الزيادة السكانية على سطح المريخ بما يفوق قدرة الكوكب، قد يعتقد خطأً أنه لا يهتم بالسلامة في الذكاء الاصطناعي. وفي الواقع، أعرف شخصيّاً أنه يهتم بذلك -النقطة الأساسية هي أنه لما كانت تقديراته الزمنية هي أطول مدة، فإنه بطبيعة الحال يميل إلى تفضيل إعطاء الأولوية لتحديات الذكاء الاصطناعي على المدى القصير منها على المدى الطويل.

خرافات حول ماهية المخاطر

قلبت عيني دهشة عندما شاهدت هذا العنوان في صحيفة **دايلي ميل**:[3] «ستيفن هوكينغ يحذر من أن صعود الروبوتات قد يكون كارثيّاً على البشر». فقد توقفت عن حساب عدد المقالات المشابهة التي رأيتها، وعادة ما يرافقها صورة روبوت شرير المظهر يحمل سلاحاً والإيحاء بأننا يجب أن نقلق من صعود الروبوتات وقتلها لنا؛ لأنها ستصبح ذات وعي و/ أو شريرة. ومن ناحية أخرى، فإن مثل هذه المقالات مثيرة للإعجاب بالفعل؛ لأنها تلخص السيناريو الذي **لا** يثير قلق زملائي العاملين في الذكاء الاصطناعي. ويجمع هذا **السيناريو ما بين ما لا يقل عن ثلاثة مفاهيم** خاطئة مستقلة: القلق بخصوص **الوعي** و**النزعة الشريرة** و**الروبوتات**، على التوالي.

وإذا كنت تقود سيارتك على الطريق؛ فلديك تجربة ذاتية Subjective experience بالألوان والأصوات، وما إلى ذلك مما يحيط بك. ولكن هل للسيارة ذاتية القيادة تجربة ذاتية؟ هل تشعر بأي شيء على الإطلاق من كونها سيارة ذاتية القيادة، أم أنها مثل زومبي لاواعٍ دون أي تجربة ذاتية؟ وعلى الرغم من أن هذا الغموض المحيط بالوعي هو أمر مثير للاهتمام في حد ذاته، وسنخصص له الفصل الثامن، بيد أنه غير ذي صلة بمخاطر الذكاء الاصطناعي، فإذا صطدمتك سيارة من دون سائق، فلن يكون هناك فرق بالنسبة إليك ما إذا كانت السيارة واعية بذاتها أم لا. وبالمثل، ما سيؤثر فينا نحن البشر هو **ما يفعله** الذكاء الاصطناعي الفائق، وليس كيف يشعر ذاتيّاً.

الخوف من أن تتحول آلات إلى كيانات شريرة هو بدوره مغالطة أخرى. والقلق الحقيقي ليس من أن تغدو خبيثة، ولكن من مدى الكفاءة. فالذكاء الاصطناعي الفائق

الذكاء Superintelligent AI من حيث التعريف جيد جدّاً في تحقيق أهدافه، مهما كانت تلك الأهداف، لذا نحتاج إلى التأكد من أن أهدافه متوائمة (متوافقة) مع أهدافنا. ففي الغالب أنتَ لست كارهاً للنمل ولا تدوسها من منطلق الخبث. ولكن إذا كنت مسؤولاً عن مشروع الطاقة الكهرمائية الخضراء وهناك جحر نمل في المنطقة التي ستغمرها المياه، فذلك من سوء حظ النمل. وما تريده حركة الذكاء الاصطناعي المفيد تجنب وضع الإنسانية في موقع هؤلاء النمل.

ويرتبط سوء فهم معنى الوعي بخرافة مفادها أنه لا يمكن أن تكون للآلات أهداف، لكن من الواضح أن الآلات قد تكون لها -بالمعنى الضيق- أهداف حتى تتمكن من إظهار سلوك موجَّه نحو هدف: فأبسط تفسير لسلوك صاروخ مُوجَّه بالحرارة هو أنه يهدف إلى قصف هدف. وإذا كنت تشعر بأنك مهدد من قبل آلة لا تتواءم أهدافها مع أهدافك، فإن أهدافها بهذا المعنى الضيق بالضبط هي التي تزعجك، وليس ما إذا كانت الآلة واعية وتشعر بإحساس وجود هدف، إذا كان هذا الصاروخ الموجه بالحرارة يطاردك، فربما لن تصرخ: «لست قلقاً؛ لأن الآلات لا يمكن أن تكون لها أهداف!»

إنني أتعاطف مع رودني بروكس Rodney Brooks وغيره من روّاد الروبوتات الذين ينزعجون من تصوير الصحف الصفراء للروبوتات على أنها شياطين مارقة؛ لأن بعض الصحافيين يبدون مهووسين بالروبوتات ويزينون العديد من مقالاتهم بصور وحوش معدنية شريرة ذات عيون حمراء لامعة. وفي الواقع، فإن حركة الذكاء الاصطناعي المفيد ليست قلقة من الروبوتات ولكن من الذكاء نفسه: على الخصوص الذكاء الذي لا تتواءم أهدافه مع أهدافنا. وما يزعجنا هو أن مثل عدم التواؤم هذا ليس بحاجة إلى روبوت فعلي، بل مجرد اتصال بالإنترنت - وسنستكشف في الفصل الرابع الكيفية التي يمكن بها أن يمكّن ذلك من التفوق على الأسواق المالية، والتفوق على ابتكارات الباحثين من البشر، والتلاعب في قادة البشر، وتطوير أسلحة لا يمكننا حتى فهم ماهيتها. وحتى لو كان بناء الروبوتات مستحيلاً فعليّاً، يمكن للذكاء الاصطناعي فائق الذكاء وفائق الغنى أن يدفع بسهولة عدداً لا يحصى من البشر -أو يتلاعب بهم- للقيام بما يريد، كما في رواية الخيال العلمي **نيورومانسر** ***Neuromancer*** من تأليف ويليام غيبسون William Gibson.

وترتبط الفكرة الخاطئة بماهية الروبوت بأسطورة أن الآلات لا تستطيع التحكم في البشر، إذ يُمكّن الذكاء من التحكم: تسود سيطرة البشر ليس لأننا أقوى، ولكن لأننا أكثر ذكاء، هذا يعني أنه إذا تخلينا عن موقفنا كأذكى ما هو على كوكبنا، فمن الممكن أن نتخلى عن السيطرة أيضاً.

يلخص الشكل 5.1 جميع هذه المفاهيم الخاطئة الشائعة، بحيث يمكننا الاستغناء عنها دفعة واحدة وإلى الأبد، ونركز في نقاشنا مع الأصدقاء والزملاء على العديد من الموضوعات الجدلية المعقولة - والتي، كما سنرى، لا يوجد نقص فيما هو مطروح منها!

الطريق إلى الأمام

وفي بقية هذا الكتاب سنستكشف معاً مستقبل الحياة مع الذكاء الاصطناعي، دعنا نبحر في هذا الموضوع الغني والمتعدد الأوجه بطريقة منظمة من خلال استكشاف القصة الكاملة للحياة استكشافاً منظّماً وعبر تسلسل زمني، ومن ثم نستكشف الأهداف والمعنى والإجراءات التي يجب اتخاذها لبناء المستقبل الذي نريده.

في الفصل الثاني، نستكشف أسس الذكاء وكيف يمكن إعادة ترتيب المادة الغبية -ظاهريّاً- لتتذكر وتحوسب وتتعلم. وبينما نحن نمضي قدماً في المستقبل، تتفرع قصتنا إلى العديد من السيناريوهات المُحدَّدة وفقا للإجابات عن بعض الأسئلة الأساسية. ويلخص الشكل 6.1 الأسئلة الرئيسية التي سنتطرق إليها ونحن نمضي إلى الأمام في المسار الزمني نحو ذكاء اصطناعي أكثر تقدماً من أي وقت مضى.

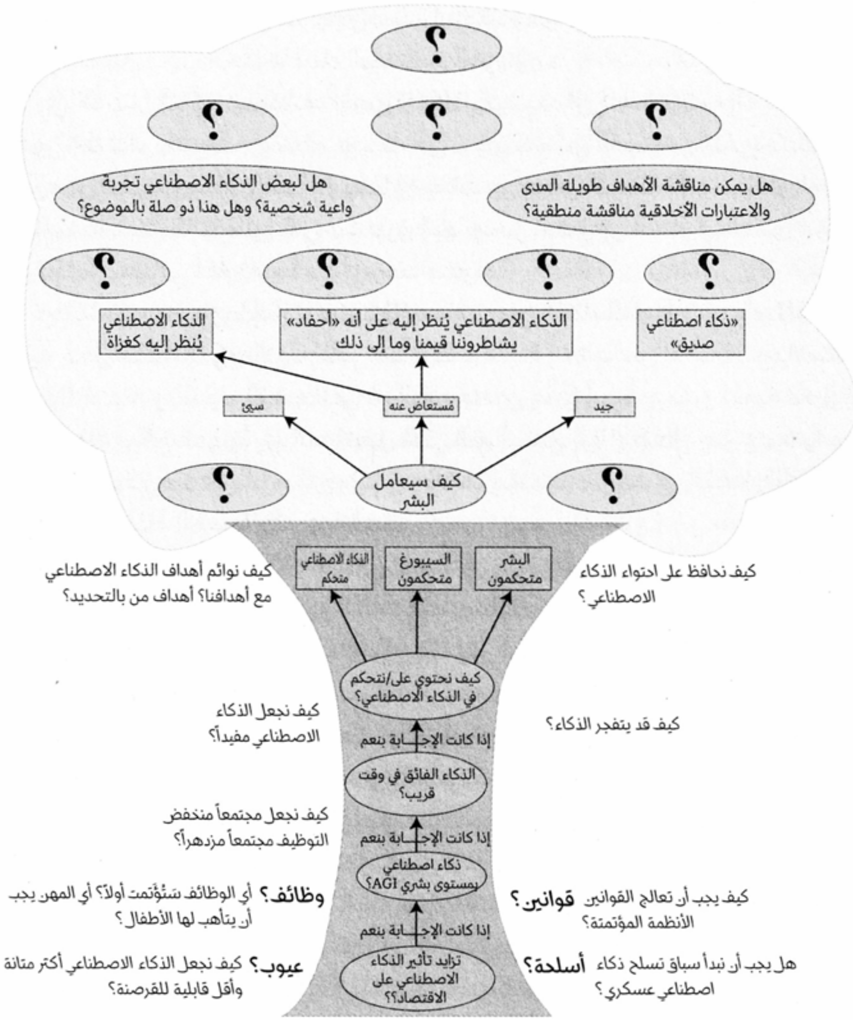
حاليا، نواجه خيار ما إذا كنا سنستهل سباق تسلح الذكاء الاصطناعي، وأسئلة حول كيفية جعل أنظمة الذكاء الاصطناعي غداً أنظمة متينة وخالية من الأخطاء. وإذا استمر التأثير الاقتصادي للذكاء الاصطناعي في النمو، يتعين علينا أيضاً أن نقرر كيفية تحديث قوانيننا والنصائح المهنية في المستقبل الآتي التي تُسدى إلى الأطفال حتى يتمكنوا من تجنب المهن التي سرعان ما ستتحول إلى وظائف مؤتمتة. ونبحث في الفصل الثالث في مثل هذه الأسئلة المعنية بالمدى القصير.

إذا استمر تقدّم الذكاء الاصطناعي وصولاً إلى المستويات البشرية؛ فعلينا أيضاً أن نسأل أنفسنا: كيف نضمن أن يكون ذلك مفيداً، وما إذا كان بإمكاننا أو يتعين علينا إنشاء مجتمع رفاهية يزدهر من دون وظائف؟ وهذا يثير أيضاً مسألة ما إذا كان تفجر الذكاء الاصطناعي أو النمو البطيء ولكن المستمر سيدفع بذكاء اصطناعي على المستوى البشري (الذكاء AGI) إلى ما هو أبعد من المستويات البشرية. أما في الفصل الرابع؛ فنستكشف مجموعة واسعة من مثل هذه السيناريوهات. ونبحث في نطاق إمكانات ما بعد الكارثة في الفصل الخامس، فيما يتراوح من مجتمع يمكن القول إنه لاطوباوي Dystopic إلى مجتمع يمكن القول إنه طوباوي. من هو المسؤول - البشر أم الذكاء الاصطناعي أم السيبورغ Cyborgs؟ هل سيُعامل البشرُ معاملة حسنة أم سيئة؟ هل تم استبدالنا، وإذا كان الأمر كذلك، فهل نعتبر بدائلنا غزاة أو أحفادا ذوي قيمة؟ أنا متشوق جدّاً لمعرفة أي سيناريوهات الفصل الخامس هي التي تفضلها أنت شخصيّاً! فقد أعددت موقعا على الويب http://AgeOfAi.org، إذ يمكنك المشاركة بآرائك والانضمام إلى المحادثة.

أخيرا، في الفصل السادس نضرب قدماً لبلايين السنين في المستقبل القادم؛ مما يمكننا من استخلاص استنتاجات أقوى مما كان بإمكاننا في الفصول السابقة، إذ إن الحدود القصوى للحياة على كوكبنا لا تتحدد بالذكاء بل بقوانين الفيزياء.

بعد الانتهاء من استكشافنا لتاريخ الذكاء، سنكرس ما تبقى من الكتاب للنظر في المستقبل الذي نهدف إليه وكيفية الوصول إليه، لنكون قادرين على ربط الحقائق المجردة

بالسؤال عن الغرض والمعنى. وسنستكشف الأساس الفيزيائي للأهداف في الفصل السابع والوعي في الفصل الثامن. وأخيراً، في الخاتمة، نستكشف ما يمكن عمله الآن للمساعدة على خلق المستقبل الذي نريده.



الشكل 6.1: أي أسئلة الذكاء الاصطناعي هي أسئلة مهمة يعتمد على أي مدى سيتطور الذكاء الاصطناعي، وأي فرع من فروع مستقبلنا سيسلك.

إذا كنت قارئاً يحب تخطي الأجزاء، فإن معظم الفصول مكتفية ذاتيّاً تماماً بمجرد استيعابك للمصطلحات والتعاريف من هذا الفصل الأول. وإذا كنت باحثاً في مجال الذكاء الاصطناعي، يمكنك اختيار تخطي مجمل الفصل الثاني باستثناء التعريفات المبدئية

للذكاء الواردة فيه. وإذا كنت مستجدّاً في مجال الذكاء الاصطناعي، فسيقدم لك الفصلان الثاني و الثالث الحجج التي تفسر سبب عدم إمكانك تجاهل الفصول من الرابع إلى السادس باعتبارها مادة علمية خيالية مستحيلة، يلخص الشكل 7.1 أين تقع الفصول المختلفة على الطيف من الحقيقة إلى الخيال.

رحلة رائعة تنتظرنا، هيا بنا نبدأ!

		العنوان المختصر للفصل	الموضوع	الحالة
		تمهيد: حكاية فريق أوميغا	غذاء للتفكير	تخميني جداً
تاريخ الذكاء	1	الحوار	مفاهيم رئيسية ومصطلحات	ليس تخمينيّاً جدّاً
	2	المادة تصبح ذكية	مبادئ الذكاء	
	3	الذكاء الاصطناعي والأسلحة والقانون	المستقبل القريب	
	4	تفجر الذكاء؟	سيناريوهات الذكاء الفائق	تخميني جداً
	5	ما بعد الحدث	العشرة آلاف سنة التالية	
	6	وقفيتنا الكوزموسية	بلايين السنوات التالية	
تاريخ المعنى	7	أهداف	تاريخ السلوك المُوجَّه بالأهداف	ليس تخمينيّاً جدّاً
	8	وعي	الوعي الطبيعي والصناعي	تخميني
		خاتمة	ما الذي يتعين علينا فعله؟	ليس تخمينيّاً جدّاً

الشكل 7.1: هيكل الكتاب

الخلاصة:

- الحياة، المُعرَّفة كعملية يمكنها الحفاظ على تعقيدها وتكرار نفسها، بإمكانها أن تتطور من خلال ثلاث مراحل: مرحلة بيولوجية (1.0)، إذ تتطور معداتها وبرمجياتها، ومرحلة ثقافية (2.0)، إذ يمكنها تصميم برمجياتها بنفسها (من خلال التعلّم)، ومرحلة تكنولوجية (3.0)، إذ يمكنها تصميم أجهزتها أيضاً، لتكون سيدة مصيرها.
- الذكاء الاصطناعي ربما يمكننا من إطلاق مرحلة الحياة 3.0 في هذا القرن، وقد نشأ حوار رائع حول المستقبل الذي يجب أن نهدف إليه وكيف يمكن تحقيق ذلك. وفي الجدل الدائر هناك ثلاثة معسكرات رئيسية: المتشككون في التقنية، واليوتوبيون الرقميون، وحركة الذكاء الاصطناعي المفيد.
- يشك المتشككون في التقنية بإمكانية تطوير ذكاء اصطناعي على المستوى البشري (الذكاء AGI)؛ ولأنه من الصعوبة بمكان فلن يحدث ذلك لمئات السنين؛ مما يجعل القلق بشأنه (وبشأن الحياة 3.0) الآن أمراً سخيفاً.

- يرى اليوتوبيون الرقميون أن ذلك من المرجح أن يحدث في هذا القرن ويرحبون بكل سعة صدر بالحياة 3.0، وينظرون إليها كخطوة طبيعية ومرغوب فيها بتطور الكوزموس.
- حركة الذكاء الاصطناعي المفيد أيضاً ترى أن حدوث ذلك في هذا القرن هو أمر محتمل، لكنهم يعتقدون أن التوصل إلى نتيجة جيدة ليس بالأمر المضمون، ولكنه أمر يحتاج إلى ضمانة من خلال العمل الجاد من خلال أبحاث سلامة الذكاء الاصطناعي.
- فيما عدا هذه الخلافات المنطقية التي يتجادل حولها أبرز خبراء العالم، هناك أيضاً نزاعات زائفة مملة بسبب سوء الفهم. مثلاً، لا تضيع وقتاً في الجدل حول "الحياة" أو "الذكاء" أو "الوعي" قبل التأكد من أنك ومُحاوِرك تستخدمان هذه الكلمات لتعني الشيء نفسه! وهذا الكتاب يستخدم التعاريف المدرجة في الجدول 1.1.
- أيضاً انتبه للمفاهيم الخاطئة الشائعة المدرجة في الشكل 5.1: «الذكاء الفائق بحلول 2100 هو محتوم/مستحيل»، «فقط اللاضيون (أعداء الآلة) يقلقون من الذكاء الاصطناعي»، «القلق هو من إمكان تحول الذكاء الاصطناعي إلى كيان شرير/ أو واعٍ، وهو أمر سيقع بعد سنوات فقط»، «الروبوتات هي مصدر القلق الرئيسي»، «الذكاء الاصطناعي ليس بإمكانه السيطرة على البشر ولا يمكن أن تكون له أهداف».
- في الفصول من الثاني إلى السادس، سنستكشف قصة الذكاء من بدايتها المتواضعة منذ بلايين السنين إلى مستقبل كوزموسي مُحتمل بعد بلايين السنين من الآن. وسنبحث في التحديات على المدى القريب مثل الوظائف، وأسلحة الذكاء الاصطناعي، والبحث عن ذكاء اصطناعي على المستوى البشري (الذكاء AGI)، ثم سنستكشف إمكانات طيف رائع من العقود المستقبلية المحتملة مع الآلات الذكية و/أو البشر، أتساءل ما هي الخيارات التي تفضلها!
- في الفصول من السابع إلى الثامن، سننتقل من الأوصاف الوقائعية الجامدة إلى استكشاف الأهداف والوعي والمعنى، والبحث فيما يمكننا فعله الآن للمساعدة على خلق المستقبل الذي نريده.
- انظر إلى هذا الحوار حول مستقبل الحياة مع الذكاء الاصطناعي باعتباره أهم حوار في زمننا هذا - يرجى الانضمام إليه!

الفصل 2

المادة تصبح ذكية

«الهيدروجين... إذا مُنِح وقتاً كافياً؛ فسيتحول إلى أشخاص».
إدوارد روبرت هاريسون، 1995.

أحد أكثر التطورات إثارة خلال 13.8 بليون سنة منذ الانفجار الكبير (العظيم) Big Bang هو أن المادة الغبية وغير المفعمة بالحياة تحولت إلى ذكاء. كيف يمكن أن يحدث هذا وكيف يمكن لأشياء أكثر ذكاء أن تتحقق في المستقبل؟ ماذا يقول العلمُ عن تاريخ ومصير الذكاء في الكوزموس؟ لمساعدتنا على الإجابة عن هذه الأسئلة، دعونا نخصص هذا الفصل لاستكشاف الأسس والكيانات الأساسية للذكاء، ماذا يعني أن نقول إن كمية مهملة من المادة ذكية؟ ماذا يعني Bن تقول إن كائناً يمكنه أن يتذكر ويحسب ويتعلم؟

ما هو الذكاء؟

مؤخراً سعدت أنا وزوجتي بحضور ندوة حول الذكاء الاصطناعي نظمتها مؤسسة نوبل السويدية، وعندما طُلب إلى لجنة من الباحثين البارزين في مجال الذكاء الاصطناعي تعريف الذكاء، جادلوا باستفاضة دون التوصل إلى توافق في الآراء، فقد وجدنا هذا مُضحكاً جدّاً: لا يوجد اتفاق على الذكاء الذي يتم حتى بين الباحثين الأذكياء المشتغلين في الذكاء! إذاً، من الواضح أنه لا يوجد تعريف «صحيح» «غير مُعارَض» للذكاء، بدلاً من ذلك، هناك العديد من التعاريف المتنافسة، بما في ذلك القدرة على التفكير المنطقي، والفهم، والتخطيط، والمعرفة العاطفية، والوعي بالذات، والإبداع، وحل المشكلات والتعلم.

في استكشافنا لمستقبل الذكاء، نود أن نتبنى رؤية واسعة وشاملة إلى حد بعيد، لا تقتصر على أنواع الذكاء الموجودة حتى الآن، وهذا هو السبب في أن التعريف الذي قدمته في الفصل السابق، والطريقة التي سأستخدم بها الكلمة في هذا الكتاب، واسعة جدّاً:

الذكاء = القدرة على تحقيق الأهداف المعقدة

هذا واسع بما يكفي ليشمل جميع التعريفات المذكورة آنفاً؛ لأن التفاهم، والوعي بالذات، وحل المشكلات، والتعلم، وما إلى ذلك، كلها أمثلة على الأهداف المعقدة التي قد تكون لدى المرء، كما أنه تعريف واسع بما يكفي لاستخدام تعريف في **قاموس أكسفورد**: «القدرة على اكتساب المعرفة والمهارات وتطبيقها» - نظراً لأن هدف المرء قد يكون تطبيق المعرفة والمهارات.

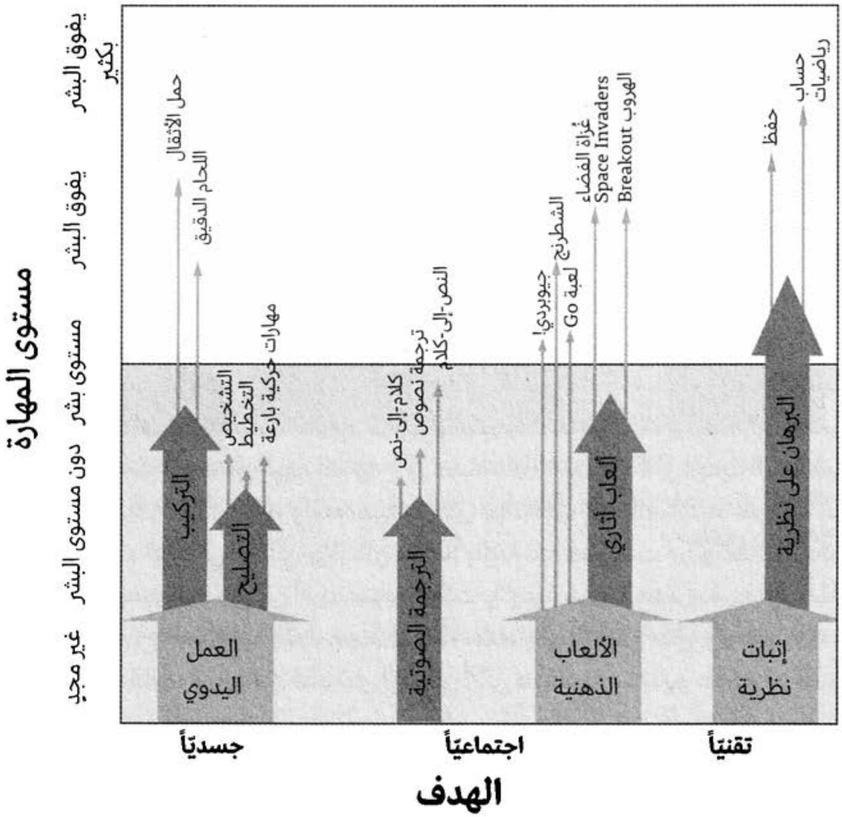
ولما كان هناك العديد من الأهداف المحتملة، فإن هناك العديد من أنواع الذكاء المُمكنة، وبناء على تعريفنا، لا معنى لقياس ذكاء البشر أو الحيوانات أو الآلات غير البشرية باستخدام رقم واحد مثل معدل الذكاء IQ،* فما هو أكثر ذكاء: برنامج حاسوب يمكنه لعب الشطرنج فقط أو لعبة يمكن أن تلعب لعبة Go فقط؟ لا توجد إجابة منطقية عن هذا؛ نظرا لأنها تجيد فعل أشياء مختلفة عن بعضها البعض ولا يمكن مقارنتها ببعضها مباشرة، ولكن، يمكننا القول إن برنامجاً ثالثاً أكثر ذكاء من كلا البرنامجين الآخرين إذا كان على الأقل جيداً في تحقيق **جميع** الأهداف، وبصورة أفضل في برنامج واحد على الأقل (الفوز في لعبة الشطرنج، لنقل).

ليس من المنطقي أيضاً أن نتجادل حول ما إذا كان شيء ما أو لا يكون ذكيّاً أم لا في الحالات الواقعة على الحدود بين الأشياء؛ نظراً لأن القدرات تأتي على شكل طيف Spectrum وليس بالضرورة على شكل القدرة على القيام بكل شيء أو لا شيء منه، فمن هم الأفراد القادرون على تحقيق هدف التحدث؟ الأطفال حديثو الولادة؟ لا، مذيعو راديو؟ نعم، ولكن ماذا عن الأطفال الذين بدؤوا المشي لتوهم، والذين يمكنهم التحدث بعشر كلمات؟ أو 500 كلمة؟ أين سترسم هذا الخط الفاصل؟ لقد استخدمت كلمة «معقد» المبهمة عن قصد في التعريف آنفاً؛ لأن محاولة رسم خط اصطناعي بين الذكي وعدم الذكي ليست بالأمر المثير للاهتمام، ومن المفيد أكثر تحديد درجة القدرة على تحقيق أهداف مختلفة.

لتصنيف الذكاءات المختلفة في نظام تصنيفي، هناك تمييز مهم آخر هو بين الذكاء الضيق والواسع، وكان جهاز الحاسوب «ديب بلو» التابع لشركة «آي بي إم»، والذي أطاح ببطل العالم في الشطرنج غاري كاسباروف عن عرشه في عام 1997، قادراً على إنجاز المهمة

* لمعرفة ذلك، تخيل كيف ستستجيب إذا ادعى شخص ما أنه يمكن تحديد مقدار القدرة على إنجاز الأعمال الرياضية على المستوى الأولمبي من خلال رقم واحد يسمى «الحاصل الرياضي» أو AQ للاختصار، بحيث يفوز اللاعب الأولمبي ذو أعلى AQ بالميداليات الذهبية في جميع الألعاب الرياضية.

الضيقة (المُحدّدة) جدّاً المتمثلة في لعب الشطرنج. وعلى الرغم من أجهزته وبرمجياته المثيرة للإعجاب، فإنه لم يتمكن حتى من التغلب على طفل في الرابعة من العمر في لعبة tic-tac-toe، يمكن لنظام DQN AI من غوغل ديب مايند Google DeepMind تحقيق نطاق أوسع قليلاً من الأهداف: يمكن للنظام أن يلعب العشرات من ألعاب الحاسوب أتاري المختلفة بمهارة



الشكل 1.2: لا يمكن قياس الذكاء، الذي يُعرّف بأنه القدرة على تحقيق الأهداف المعقدة، برقم واحد هو معدل ذكاء IQ، بل يُقاس باستخدام طيف من القدرة عبر جميع الأهداف. يشير كل سهم إلى مدى كفاءة أفضل أنظمة الذكاء الاصطناعي اليوم في تحقيق أهداف متنوعة؛ مما يوضح أن الذكاء الاصطناعي يميل إلى أن يكون **ضيّقاً**، وكل نظام منه قادر على تحقيق أهداف محددة. وعلى النقيض من ذلك، فإن الذكاء البشري واسع بشكل ملحوظ: فطفل سليم سيتعلم الكيفية التي سيصبح بها أفضل في أي شيء تقريباً.

تصل إلى المستوى البشري أو أفضل، وعلى النقيض من ذلك، فإن الذكاء البشري واسع إلى حد بعيد ولا يوجد له نظير حتى الآن، وقادر على إتقان مجموعة مذهلة من المهارات، فطفل سليم يحصل على وقت كافٍ من التدريب قادر على أن يجيد ليس فقط أي لعبة، ولكن أيضاً في أي لغة أو رياضة أو مهنة، وعند مقارنة ذكاء البشر والآلات اليوم، نجد أننا نحن البشر ننتصر باكتساح، في حين أن الآلات تتفوق علينا في أداء عدد صغير ولكن متزايد

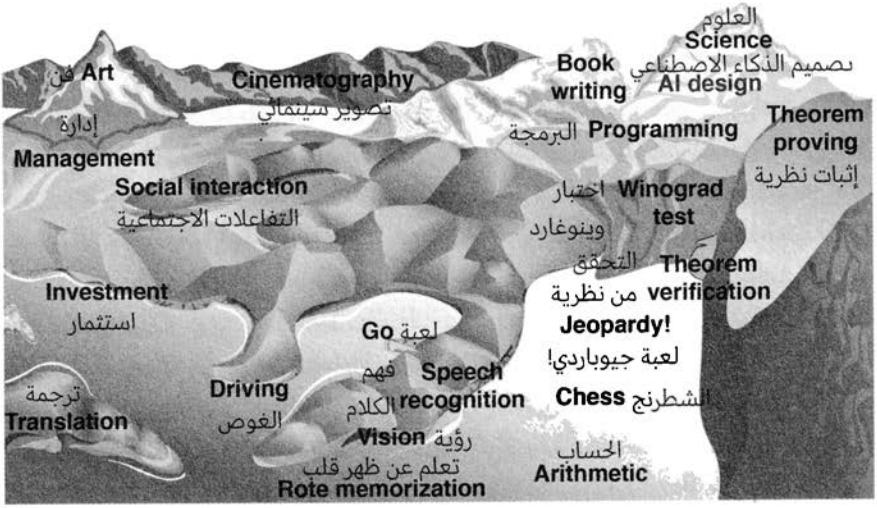
من المجالات الضيقة، كما هو موضح في الشكل 1.2، والكأس المقدسة التي تسعى إليها أبحاث الذكاء الاصطناعي هي بناء «ذكاء اصطناعي عام» (المعروف بالذكاء الاصطناعي العام، AGI) يكون واسع النطاق إلى أقصى حد: أي قادراً على تحقيق أي هدف تقريباً، بما في ذلك التعلم، وسنستكشف ذلك بالتفصيل في الفصل الرابع. وقد جعل باحثو الذكاء الاصطناعي شين ليغ Shane Legg ومارك غبرود Mark Gubrud وبين غورتزيل Ben Goertzel مصطلح الذكاء «AGI» (الذكاء الاصطناعي العام، Artificial general intelligence) مصطلحاً شائعاً ليعني بشكل أكثر تحديداً الذكاء الاصطناعي على **مستوى الإنسان**: القدرة على تحقيق أي هدف على الأقل بنفس كفاءة البشر على الأقل.[1] وسألتزم بتعريفهم، لذا ما لم أضع قيمة للمصطلح (من خلال كتابة ذكاء يفوق البشر AGI مثلاً)، سأستخدم مصطلح الذكاء «AGI» كاختزال لـ «ذكاء على مستوى البشر AGI».*

على الرغم من أن كلمة «ذكاء» تميل إلى أن تكون لها دلالات إيجابية، فإن من المهم ملاحظة أننا نستخدمها بطريقة محايدة تماماً: كقدرة على تحقيق أهداف معقدة بغض النظر عما إذا كانت هذه الأهداف جيدة أم سيئة. ومن ثم، قد يكون الشخص الذكي جيداً جدّاً في مساعدة الأشخاص أو تحقيق نتائج جيدة جدّاً في إيذاء الأشخاص، وسنستكشف موضوع الأهداف في الفصل السابع. ففيما يتعلق بالأهداف، نحتاج أيضاً إلى توضيح مدى دقة أهدافنا التي نشير إليها، لنفترض أن مساعدك الشخصي الآلي الجديد المستقبلي ليست له أهداف من أي نوع، ولكنه سيفعل كل ما تطلب إليه القيام به، فتطلب إليه تحضير العشاء الإيطالي المثالي، وإذا كان متصلاً بالإنترنت؛ فسيبحث عن وصفات العشاء الإيطالية، وكيفية الوصول إلى أقرب سوبر ماركت، وكيفية سلق المعكرونة، وما إلى ذلك، ثم يشتري بنجاح المكونات ويُعدُّ وجبة لذيذة. ستفترض أنه ذكي حتى ولو كان الهدف الأصلي هو هدفك. وفي الواقع، فقد تبنى الإنسان الآلي هدفك بعد تقديم طلبك، ثم قسمه إلى أهداف فرعية لنفسه تتوالى في تسلسل هرمي، من دفع النقود لأمين الصندوق إلى برش جبن البارميجان. بهذا المعنى، يرتبط السلوك الذكي ارتباطاً وثيقاً بتحقيق الأهداف.

بالنسبة إلينا، فإن من الطبيعي تقييم صعوبة المهام بناء على مدى صعوبة أدائها من قبلنا نحن البشر، كما هو موضح في الشكل 1.2، ولكن هذا قد يعكس صورة مضللة عن مدى الصعوبة بالنسبة إلى أجهزة الحاسوب، فقد يبدو لنا أن ضَرْب 314,159 في 278,828 أصعب من التعرف على صديق في صورة، لكن الحواسيب هزمتنا في الحساب قبل فترة طويلة من ولادتي، في حين أن التعرف على الصور على مستوى الإنسان لم يصبح ممكناً إلا مؤخراً، وهذه الحقيقة -من أن المهام الحسية ذات المستوى المنخفض تبدو سهلة على الرغم من أنها تتطلب موارد حاسوبية هائلة- تُعرف

* يفضل البعض مصطلح «الذكاء الاصطناعي على المستوى البشري» أو «الذكاء الاصطناعي القوي» كمرادفين للذكاء AGI (الذكاء الاصطناعي العام Artificial general intelligence)، لكن كليهما يمثل مشكلة، حتى حاسبة الجيب هي عبارة عن آلة ذكية على المستوى البشري بالمعنى الضيق. إن المعنى المناقض لمصطلح «ذكاء اصطناعي قوي» هو «ذكاء اصطناعي ضعيف»، ولكن يبدو من الغريب أن نطلق على أنظمة الذكاء الاصطناعي الضيقة من مثل ديب بلو Deep Blue وواتسون Watson وآلفاغو AlphaGo وصف أنظمة«ضعيفة».

بمفارقة مورافيك Moravec's paradox، وتفسر حقيقة أن دماغنا يجعل هذه المهام تبدو سهلة من خلال تكريس كميات هائلة من الأجهزة المخصصة لها - أكثر من ربع أدمغتنا، في الواقع.



الشكل 2.2: رسم توضيحي لمصطلح هانز مورافيك «التضاريس الطبيعية للكفاءة البشرية»، إذ يمثل الارتفاعُ مدى صعوبة الإنجاز بالنسبة إلى الحواسيب، ويمثل مستوى ارتفاع سطح البحر ما تستطيع الحواسيب فعله.

وأنا أحب هذا التشبيه الذي وضعه هانز مورافيك، وقد استخدمته لإعداد الشكل 2.2:

> أجهزة الحاسوب هي آلات عالمية، إمكاناتها تمتد بشكل منتظم عبر نطاق غير محدود من المهام، ومن ناحية أخرى، تتمتع الإمكانات البشرية بالقوة في الأمور التي كانت منذ أمد ذات أهمية بالغة للبقاء على قيد الحياة، ولكنها ضعيفة في الأشياء غير ذات الصلة ذلك، تخيَّلْ «التضاريس الطبيعية للكفاءة البشرية»، بأراضٍ منخفضة يُطلق عليها تسميات مثل «الحساب» و«الحفظ عن ظهر قلب»، وتلال مثل «إثبات نظرية» و«لعب الشطرنج»، وقمم جبال عالية مثل «الحركة» و«التنسيق اليد والعين» و«التفاعل الاجتماعي»، وتطور أداء الحاسوب يشبه الماء الذي يفيض على التضاريس الطبيعية ببطء. وقبل نصف قرن بدأت المياه بإغراق الأراضي المنخفضة، وإنهاء دور الآلات الحاسبة البشرية والكَتَبَة، ولكن ترك على أراضٍ مرتفعة وجافة، والآن وصل الفيضان إلى سفوح التلال، ومراكزنا الحدودية هناك تفكر في التقهقر، وما زلنا نشعر بالأمان على قممنا، ولكن بالمعدل الحالي، ستُغمر أيضاً خلال نصف قرن آخر، أقترح أن نبني سفينة مع اقتراب ذلك اليوم، وتبني حياة بحرية![2]

خلال العقود التي مرت منذ كتابة هذه الفقرات، استمر مستوى سطح البحر بالارتفاع بلا هوادة، كما تنبأ مورافيك، كما لو كان الاحترار العالمي يتعاطى المنشطات الستيرويدية، وبعض من السفوح (بما في ذلك لعبة الشطرنج) غارقة منذ فترة طويلة، ما يأتي بعد ذلك وما يجب أن نفعله حيال الموضوع هو موضوع بقية هذا الكتاب.

ومع استمرار ارتفاع مستوى سطح البحر، فقد يصل في يوم من الأيام إلى نقطة تحول؛ مما يؤدي إلى حدوث تغير جذري. ومستوى البحر الحرج هذا هو المستوى الذي تصبح عنده الآلات قادرة على تصميم ذكاء اصطناعي بنفسها، ولكن قبل الوصول إلى نقطة التحول هذه، يرتفع مستوى سطح البحر بواسطة آلات حسّنها البشر؛ وبعد ذلك، قد يكون الدافع وراءه هو آلات تُحسّن آلات، وربما أسرع بكثير مما يستطيعه البشر، فتُغمر بسرعة كلُ الأراضي. هذه هي الفكرة الرائعة والمثيرة للجدل حول **الحدث المنفرد** ***Singularity***، التي سنستمتع بها في الفصل الرابع.

وضع آلان تورنغ Alan Turing -رائد الحاسوب- إثباتاً شهيراً مفاده أنه إذا تمكن الحاسوب من إجراء مجموعة معينة من العمليات، فبإمكانه، إذا ما أعطي ما يكفي من الوقت والذاكرة، برمجةَ أي شيء يمكن أن يفعله أي حاسوب آخر، وتسمى الأجهزة التي تتجاوز هذه العتبة الحرجة حواسيب عامة Universal computers (المعروفة أيضاً بحواسيب تورينغ العالمية Turing-universal)؛ وجميع الهواتف الذكية وأجهزة الحاسوب المحمولة اليوم عالمية بهذا المعنى، على نحو مماثل، أودُّ أن أفكر في عتبة الذكاء الحرجة المطلوبة لتصميم الذكاء الاصطناعي باعتباره **عتبة الذكاء العالمي** ***Universal intelligence***: إذا ما توفّر ما يكفي من الوقت والموارد، يمكنه أن يجعل نفسه قادراً على تحقيق أي هدف وكذلك تطوير أي كيان ذكي آخر. مثلاً، إذا قرر الذكاء العالمي أن يطور مهارات اجتماعية أفضل، أو مهارات تنبؤ أفضل، أو مهارات تصميم ذكاء اصطناعي أفضل؛ فسيمكنه تحقيق ذلك. وإذا قرر معرفة كيفية بناء مصنع روبوت؛ فيمكنه فعل ذلك. وبعبارة أخرى، فإن الذكاء العالمي لديه القدرة على التطور إلى الحياة 3.0.

الحكمة التقليدية السائدة بين الباحثين في مجال الذكاء الاصطناعي هي أن الذكاء -في المُحصِّلة- هو كل ما هو متعلق بالمعلومات والحوسبة، وليس باللحم أو الدم أو ذرات الكربون، وهذا يعني أنه لا يوجد سبب أساسي لعدم قدرة الآلات في يوم من الأيام على أن تكون ذكية على الأقل.

ولكن ما هي المعلومات والحوسبة حقّاً. لما كانت الفيزياء قد علمتنا أن على المستوى الأساسي، كل شيء هو مجرد مادة وطاقة تتجولان في الأنحاء؟ كيف يمكن لشيء ما مجرّد وغير ملموس وأثيري -كالمعلومات والحوسبة- أن يُجسَّد كالأشياء المادية الملموسة؟ على وجه الخصوص، كيف يمكن لمجموعة من الجزيئات البُكم تتحرك في جميع الأنحاء وفقاً لسلوك الفيزياء القوانين أن تعرض سلوكاً بإمكاننا أن نصفه بأنه ذكي؟

إذا رأيت أن الإجابة عن هذا السؤال هي إجابة واضحة واعتقدت أنه من المعقول أن الآلات ستصبح ذكية مثل البشر في هذا القرن -مثلاً، لأنك باحث من الباحثين المشتغلين في مجال الذكاء الاصطناعي- يرجى تخطي بقية هذا الفصل والانتقال مباشرة إلى الفصل الثالث، وإن لم تكن كذلك؛ فسيسعدك أن تعرف أنني كتبت الأقسام الثلاثة التالية خصيصاً لك.

ما هي الذاكرة؟

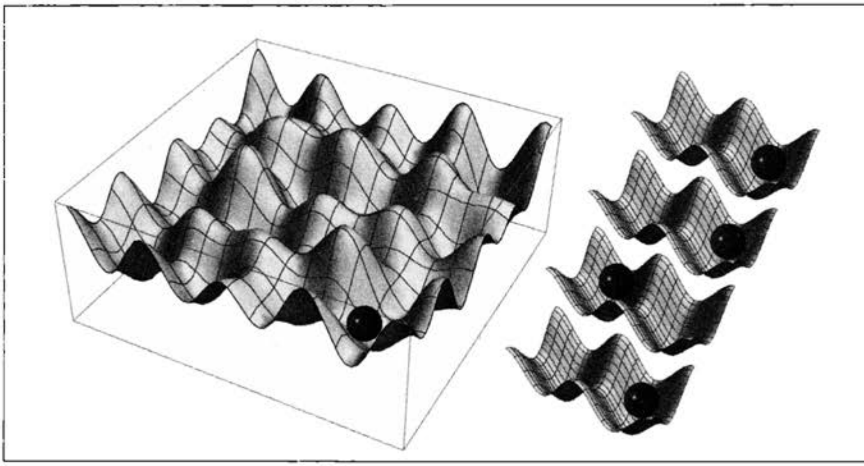
إذا قلنا إن أطلساً يحتوي على **معلومات** حول العالم، فنحن نقصد أن هناك علاقة بين حالة الكتاب (على وجه الخصوص، مواضع جزيئات معينة تعطي الحروف والصور ألوانها) وحالة العالم (مثلاً: مواقع القارات)، إذا كانت القارات في أماكن مختلفة؛ فستكون تلك الجزيئات في أماكن مختلفة أيضاً. نحن البشر نستخدم مجموعة من الأجهزة المختلفة لتخزين المعلومات، من الكتب والأدمغة إلى محركات الأقراص الصلبة، وهي جميعاً تتشارك في هذه الخاصية: أن حالتها قد ترتبط (ومن ثم تُبلغنا عن) بحالة الأشياء الأخرى التي نهتم بها.

ما هي الخصائص الفيزيائية الأساسية المشتركة بينها التي تجعلها مفيدة كأجهزة ذاكرة، أي أجهزة لتخزين المعلومات؟ الجواب هو أنها جميعاً **يمكنها أن تأتي في أشكال تستمر لفترة طويلة** بما يكفي لترميز المعلومات والاحتفاظ بها إلى وقت الحاجة. كمثال بسيط، لنفترض أنك تضع كرة على سطح أرضٍ ذات 16 وادياً، كما هو موضح في الشكل 3.2، وحينما تتدحرج الكرة وتصل إلى السكون، ستكون في أحد الأماكن 16، بحيث يمكنك استخدام موقعها كطريقة لتذكر أي رقم يتراوح بين 1 و 16.

وجهاز الذاكرة هذا متين إلى حد ما؛ لأنه حتى إذا تعرض للقليل من التشويش والإزعاج من قبل قوى خارجية؛ فمن المرجح أن تبقى الكرة في الوادي نفسه الذي وضعتَها فيه، وبإمكانك معرفة الرقم المُخزَّن. والسبب في أن هذه الذاكرة مستقرّة جدًّا هو أنّ رفع الكرة من الوادي يتطلب طاقة أكثر مما يُرجّح أن توفره الاضطرابات العشوائية. وهذه الفكرة نفسها قد توفر ذكريات مستقرة عموماً بشكل أكبر بكثير من حالة كرة متحركة: فطاقة نظام فيزيائي معقد يمكن أن تعتمد على العديد من الخواص الميكانيكية والكيميائية والكهربائية والمغناطيسية. وما دامت تستهلك الطاقة لتغيير النظام من الحالة التي تريد أن تتذكرها، فإن هذه الحالة ستكون مستقرة، وهذا هو السبب في أن المواد الصلبة لديها العديد من الحالات الطويلة العمر، ولكن ليس لدى السوائل والغازات مثل ذلك: إذا نقشت اسم شخص ما على خاتم ذهب، فإن المعلومات ستظل هناك بعد سنوات لأن إعادة تشكيل الذهب تتطلب طاقة كبيرة، ولكن إذا نقشته على سطح البِركة، فسيفقد في غضون ثانية في حين يغير سطح الماء شكله بسهولة.

أبسط جهاز الذاكرة ممكن لديه حالتان مستقرتان فقط (الشكل 3.2)، لذا يمكننا التفكير في الأمر على أنه ترميز ثنائي الأرقام Binary digit (ومنها اشتق الاختصار «بت» bit)، أي صفر أو واحد، ويمكن تخزين المعلومات المخزنة بواسطة أي جهاز ذاكرة أكثر تعقيداً في وحدات «بت» متعددة: مثلاً، إذا نظرنا إليها معاً؛ فقد تكون البتات الأربع المبينة في الشكل 3.2 (يمين) 2 × 2 × 2 × 2 = 16 حالة مختلفة: 000، 0001، 0010، 0011،....، 1111، بحيث تكون لديها مجتمعة سعة الذاكرة نفسها تماماً مثل نظام أكثر تعقيداً ذي 16 حالة (يسار)، لذا يمكننا التفكير في البتات كما لو كانت ذرات من المعلومات - أصغر جزء غير

قابل للتجزئة من المعلومات التي لا يمكن تقسيمها إلى أجزاء أخرى، والتي قد تتحد لتشكل أي معلومة. مثلاً، طبعت الكلمة Word، وقام جهاز الحاسوب المحمول الخاص بي بتمثيلها في ذاكرته كالتسلسل المكون من 4 أعداد 119 111 114 100، وتخزين كل من هذه الأعداد كثمانية بتات (يمثل الحاسوب كل حرف إنجليزي صغير بعدد يساوي 96 إضافة إلى ترتيبه في الأبجدية)، حالما أضغط على المفتاح w في لوحة المفاتيح بحاسوبي، سيعرض حاسوبي المحمول صورة مرئية لـ w على شاشتي، وتمثل هذه الصورة أيضاً بتات: 32 بتّاً تحدد لون كل من الملايين من بكسلات الشاشة.



الشكل 3.2: الكائن المادي هو جهاز ذاكرة مفيد، إذ يمكنه أن يكون في العديد من الحالات المستقرة المختلفة، فيمكن للكرة على اليسار أن تُرمِّز أربعة بتات من تصنيف المعلومات، وأي واحد منها يحتوي على $2^4 = 16$ وادياً. وتشكل الكرات الأربع الموجودة على اليمين أيضاً أربعة أجزاء من المعلومات - بت واحد لكل منها.

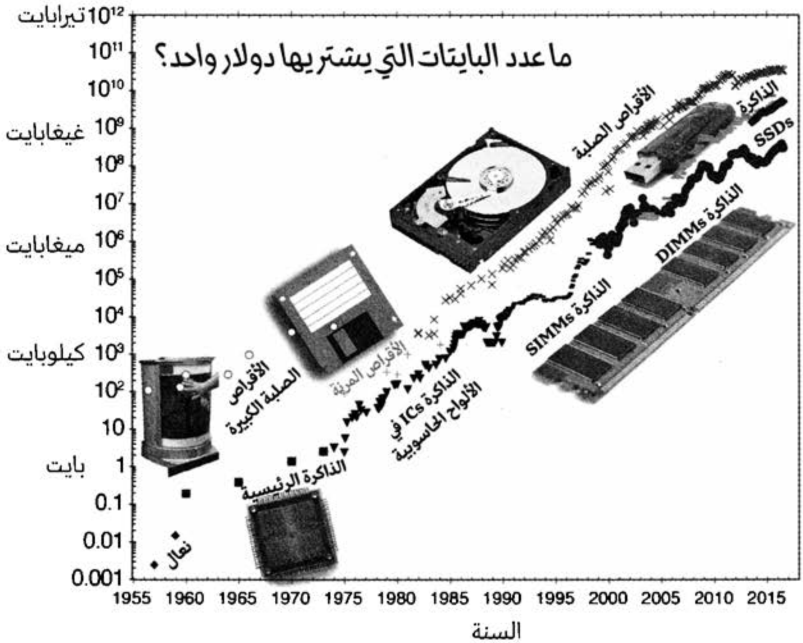
ولما كانت أنظمة الحالتين Two-state systems سهلة التصنيع والتعامل معها، فإن معظم أجهزة الحاسوب الحديثة تُخزِّن معلوماتها على هيئة بتات، ولكن هذه البتات تتجسد في مجموعة متنوعة من الطرق، فعلى قرص المدمج DVD يتوافق كل بت مع ما إذا كان هناك حفرة مجهرية أو لا عند نقطة معينة على السطح البلاستيكي، أما على القرص الصلب؛ فإن كل بت يتوافق مع نقطة على السطح مُمغنطة بإحدى طريقتين. وفي الذاكرة العاملة Working memory في حاسوبي المحمول، يتوافق كل بت مع أوضاع إلكترونات مُعيّنة، فيُحدِّد ما إذا كان الجهاز الذي يُسمى المُكثِّف الدقيق Microcapacitor مشحوناً أم لا، وبعض أنواع البتات ملائمة للنقل كذلك حتى عند سرعة الضوء: مثلاً، في الألياف الضوئية التي ترسل البريد الإلكتروني الخاص بك، كل قطعة تتوافق مع شعاع ليزر قوي أو ضعيف في وقت معين.

يفضل المهندسون ترميز (تشفير) Encode البتات في أنظمة ليست فقط مستقرة وسهلة القراءة (كخاتم من الذهب) فحسب، ولكن من السهل الكتابة عليها أيضاً: تغيير حالة القرص الثابت في حاسوبك يتطلب طاقة أقل بكثير من نقش الذهب، كما يفضلون

الأنظمة المناسبة للعمل عليها والتي يمكن إنتاجها بكميات ضخمة بتكلفة مادية منخفضة، ولكن بخلاف ذلك، ببساطة لا يهتمون في كيفية تمثيل البتات كأشياء مادية - ولا أنت في معظم الأوقات؛ لأنه لا يهم ببساطة! إذا أرسلت بريداً إلكترونيّاً إلى صديقك لطباعة مستند؛ فقد تُنسخ المعلومات في تتابع سريع من عمليات التمغنط على محرك الأقراص الثابتة إلى الشحنات الكهربائية في الذاكرة العاملة بالحاسوب، أو موجات الراديو في الشبكة اللاسلكية، أو الفولتية في جهاز التوجيه Router، أو نبضات الليزر في جهاز الألياف ضوئية، وأخيراً، جزيئات على قطعة من الورق. وبعبارة أخرى، **يمكن للمعلومات أن تسلك حياة خاصة بها، بغض النظر عن ركائزها المادية!** بالطبع، عادة ما يكون هذا الجانب المستقل من المعلومات هو الجانب الوحيد الذي نرغب فيه: إذا اتصل بك صديقك لمناقشة الوثيقة التي أرسلتها، فربما لا يتصل لمناقشة الفولتية أو الجزيئات، وهذا هو أول دليل يرشدنا إلى كيفية تجسيد شيء غير ملموس كالذكاء في أشياء مادية ملموسة، وسنرى قريباً الكيفية تكون بها فكرة استقلالية المادة هذه هي فكرة أعمق بكثير، ليس فقط فيما يتعلق بالمعلومات فحسب بل أيضاً بالحوسبة والتَّعلم.

وبسبب استقلالية المادة، تمكن المهندسون الأذكياء من الاستعاضة عن الذاكرة داخل أجهزة الحاسوب بشكل مطّرد بأخرى أفضل بشكل كبير، بناء على تقنيات جديدة ومن دون الحاجة إلى أدنى تغييرات في برمجياتنا. وكانت النتيجة مذهلة، فكما هو موضح في الشكل 4.2: على مدى العقود الستة الماضية، كان سعر ذاكرة الحاسوب يقلّ إلى نصف التكلفة كل عامين تقريباً، فقد صارت الأقراص الصلبة أرخص بمئة مليون مرة، وصارت الذاكرة الأسرع التي تفيد في الحوسبة بدلاً من مجرد التخزين أقل كلفة بمقدار عشرة تريليونات مرة. وإذا كان بإمكانك الحصول على خصم يعادل «99.9999999999 %» على جميع مشترياتك، فيمكنك شراء جميع عقارات مدينة نيويورك مقابل عشرة سنتات تقريباً وجميع الذهب المُستخرج في العالَم مقابل دولار تقريباً.

بالنسبة إلى الكثيرين منا، ترتبط التحسينات المذهلة في تكنولوجيا الذاكرة بقصص شخصية، أتذكر أنني كنت أعمل في محل لبيع الحلوى في المدرسة الثانوية لأدفع ثمن جهاز حاسوب بذاكرة تعادل 16 كيلوبايت، وعندما طوّرت برنامج معالج كلمات Word processor يعمل على هذا الحاسوب مع زميلي في المدرسة الثانوية ماغنوس بودين Magnus Bodin وبعنا البرنامج، أجبرنا على كتابة البرمجة باستخدام لغة برمجة آلية مضغوطة جدّاً لترك ذاكرة كافية للكلمات التي كان من المفترض أن يعالجها البرنامج. وبعد التعود على الأقراص المرنة Floppy drives التي تخزن 70 كيلوبايت، ذُهلت بالأقراص المرنة الأصغر حجما 3.5 بوصة. والتي قد تخزن حجماً ضخماً يصل إلى 1.44 ميغابايت وتخزّن كتابا بكامله، ثم أول قرص ثابت على الإطلاق يخزّن 10 ميغابايت -والذي لا يصلح إلا لتخزين تنزيل أغنية واحدة من الأغاني في يومنا هذا. وقبل أيام، فإن هذه الذكريات من مراهقتي كادت تبدو كما لو كانت ذكريات غير حقيقية، عندما أنفقتُ نحو 100 دولار على شراء قرص صلب بسعة أكبر بنحو 300 ألف مرة.

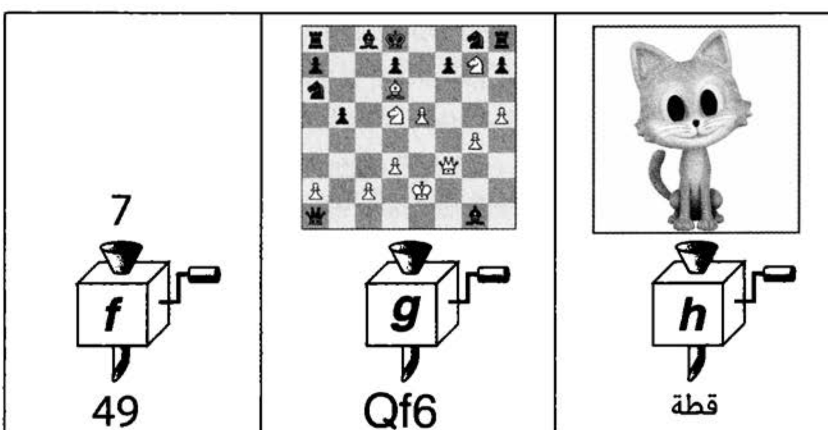


الشكل 4.2: على مدى العقود الستة الماضية انخفضت تكلفت شرائح الحاسوب إلى النصف كل سنتين تقريباً، أي ما يعادل أرخص بألف مرة كل 20 سنة تقريباً، والبايت يساوي ثمانية بتات. البيانات إهداء من جون ماكالوم، http://www.jcmi.net/memoryprice.htm.

ماذا عن أجهزة الذاكرة التي تطورت بنفسها بدلاً من أن تصمم من قبل البشر؟ علماء الأحياء لا يعرفون حتى الآن كيف كان شكل الحياة الأول من نوعه الذي نسخت الأجيال التالية مخططاته، ولكنه ربما كان صغيراً جدّاً، ففي عام 2016 صنع فريق بقيادة فيليب هوليغر Philipp Holliger من جامعة كيمبريدج جزيئاً من الحمض النووي RNA (الرنا) رمَّزَ 412 بتّاً من المعلومات الوراثية وتمكَّن من نسخ سلاسل من الحمض النووي RNA أطول أكثر من نفسه؛ مما عزز فرضية «عالمَ الحمض النووي RNA» القائلة إن حياة الأرض المبكرة اشتملت على قطع قصيرة من حمض نووي RNA قادر على نسخ نفسه.وحتى الآن، فإن أصغر جهاز ذاكرة معروف تطور عضويّاً ويُستخدم في الحياة البرية هو جينوم البكتيريا **كانديداتوس كارسونيلّا رودي** ***Candidatus Carsonella*** ruddii الذي يخزن نحو 40 كيلوبايت، في حين يُخزِّن حمضنا النووي البشري نحو 1.6 غيغابايت، أي ما يعادل تنزيل فيلم من الإنترنت، كما ذكرنا في الفصل السابق. تخزن أدمغتنا معلومات أكثر بكثير من جيناتنا: تقريباً 10 غيغابايت كهربائية (تُحدِّد أيا من المئة بليون عصبون ستقدح الإشارات العصبية في أي وقت من الأوقات) و100 تيرابايت كيميائيّاً/بيولوجيّاً (تُحدِّد مدى

قوة ارتباط الخلايا العصبية عبر المشتبكات العصبية)، تبين مقارنة هذه الأعداد بالذكريات الآلية أن أفضل أجهزة الحاسوب في العالم يمكنها الآن التفوقَ على أي نظام بيولوجي في التذكر- وبتكلفة تنخفض بسرعة وكانت تعادل بضعة آلاف من الدولارات في عام 2016.

غير أن الذاكرة في دماغك تعمل بشكل مختلف تماماً عن ذاكرة الحاسوب، ليس فقط في كيفية بنائها، ولكن أيضاً في كيفية استخدامها، فبينما تسترجع الذكريات من حاسوب أو محرك أقراص صلبة عن طريق تحديد مكان تخزينها، يمكنك استرجاع الذكريات من دماغك عن طريق تحديد شيء حول ما يتم تخزينه، وتحتوي كل مجموعة من وحدات البت في ذاكرة الحاسوب الخاص بك على عنوان رقمي، ولاستعادة جزء من المعلومات، يحدد الحاسوبُ العنوانَ الذي تريد البحث عنه، تماماً كما لو كنت أقول: «اذْهَبْ إلى رف خزانة كتبي، وتناولْ الكتاب الخامس من اليمين من على الرف العلوي، وأخبرني بما هو في الصفحة 314». وفي المقابل، يمكنك استرداد المعلومات من دماغك بطريقة مشابهة لكيفية استعادتها من محرك البحث: يمكنك تحديد جزء من المعلومات أو شيء متعلق به، وتظهر المعلومة، وإذا قلت لك «أن أكون أو لا أكون»، أو إذا كنت أبحث عنها باستخدام غوغل، فهناك احتمال أن تؤدي إلى «أن أكون، أو لا أكون، هذا هو السؤال». في الواقع، سينفع الأمر على الأرجح حتى لو استخدمت جزءاً آخر من الاقتباس أو أخطأت فيه إلى حد ما، تُعرف أنظمة الذاكرة هذه **بالترابط الذاتي** ***Auto-associative***، إذ إنها تتذكر بالارتباط بدلاً من العنوان.



الشكل 5.2: الحسابُ يأخذ المعلوماتِ ويحولها مُطبّقاً ما يسميه علماء الرياضيات دالة Function، تأخذ الدالة *f* (يسار) وحدات بت تمثل عدداً وتحسب مربعها، وتأخذ الدالة g (في الوسط) وحدات بت تمثل وضعية شطرنج وتحسب أفضل حركة للأبيض، وتأخذ الدالة h (يمين) بتات تمثل صورة وتحسب علامة نصية تصفها.

في ورقته الشهيرة عام 1982، أظهر الفيزيائي جون هوبفيلد John Hopefield الكيفية التي يمكن بها لشبكة من العصبونات المترابطة أن تعمل كذاكرة ترابطية ذاتية. أجد الفكرة الأساسية فكرة جميلة جدّاً، وهي فاعلة في أي نظام مادي بحالات مستقرة متعددة. مثلاً، فكر في كرة على سطح ذي واديين، مثل نظام بت واحد في الشكل 3.2، ودعنا نكوّن السطح بحيث تكون إحداثيات x لأدنى نقطتين، إذ يمكن للكرة أن تستقر هي $x = \sqrt{2} \approx 1.41421$ و $x = \pi \approx 3.14159$، على التوالي، إذا كنت تتذكر فقط أن π قريبة من 3، فكل ما عليك هو وضع الكرة عند النقطة إذ $x = 3$ ومشاهدتها تكشف عن قيمة أكثر دقة عندما تتدحرج إلى الحد الأدنى. أدرك هوبفيلد أن شبكة معقدة من العصبونات توفر مشهداً طبيعيّاً مشابهاً في العديد من مناطق الطاقة الدنيا التي قد يستقر عندها النظام، وأُثبِت فيما بعد أنه يمكنك ضغط ما يصل إلى 138 ذكرى مختلفة في كل ألف عصبون دون التسبب في حدوث ارتباك كبير.

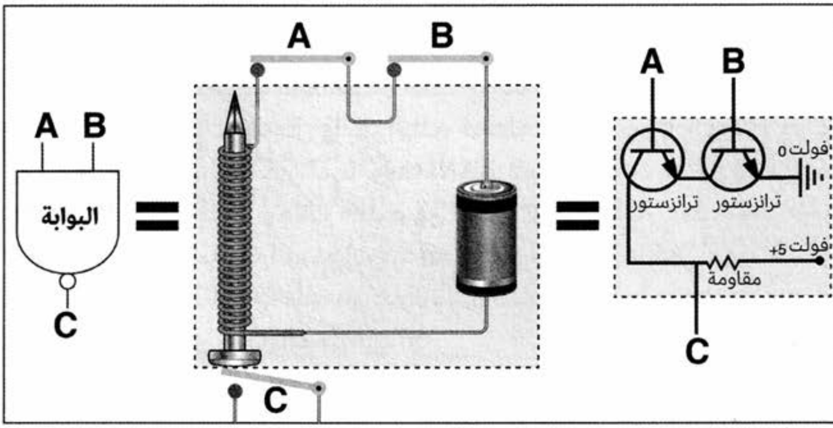
ما الحوسبة؟

رأينا الآن الكيفية التي يستطيع بها كائن مادي ما أن يتذكر المعلومات، لكن كيف يمكن أن تجري عمليات حوسبة Compute؟

الحوسبة Computation هي تحويل حالة ذاكرة Memory state واحدة إلى أخرى، بعبارة أخرى، تأخذ الحوسبة المعلومات وتُحوِّلها Transform، وتُنفِّذ الحوسبة ما يسميه علماء الرياضيات دالة Function. وأنا أنظر إلى الدالة كمفرمة لحم تطحن المعلومات، كما هو موضح في الشكل 5.2: أنت تضع المعلومات في الأعلى، وتدير ذراع المفرمة؛ فتحصل على المعلومات المُعالَجة في الأسفل - يمكنك تكرار ذلك عدة مرات كما تريد باستخدام مدخلات مختلفة. ومعالجة المعلومات هذه حتمية، بمعنى أنه إذا قمت بتكرارها بالمدخلات نفسها، فإنك تحصل على الناتج نفسه في كل مرة.

على الرغم من أنها تبدو بسيطة إلى درجة خادعة، فإن فكرة دالة هي فكرة عامة لدرجة مدهشة. وبعض الدوال بسيطة نوعاً ما، مثل الدالة التي تسمى NOT التي تُدخل بتّاً واحداً ويكون الناتج عكس المُدخل، ومن ثم تحول الصفر إلى واحد والواحد إلى صفر.وتتطابق الوظائف التي نتعلمها في المدرسة عادة مع الأزرار الموجودة على الآلة الحاسبة الجيبية، وندخل عدداً واحداً أو أكثر ونخرج عددا واحدا - مثلاً، تقوم الدالة x2 ببساطة بإدخال أعداد ومخرجات تضربها بنفسها، وهناك ظائف أخرى ربما تكون معقدة جدّاً، مثلاً، إذا كانت بحوزتك دالة تُدخل وحدات البت التي تمثل مواضع شطرنج عشوائية وبتات إخراج تمثل أفضل خطوة تالية ممكنة؛ فيمكنك استخدامها للفوز ببطولة العالم الشطرنج الحاسوبي. وإذا كنت تمتلك دالة تعمل على إدخال جميع البيانات المالية العالمية وإخراج أفضل خيارات شراء الأسهم، فقريباً ستكون غنيا جدّاً، ويكرّس العديد من الباحثين جهودهم في مجال الذكاء الاصطناعي لمعرفة كيفية تنفيذ

دوال معينة، مثلاً، الهدف من أبحاث الترجمة الآلية هو تنفيذ وحدات إدخال مدمجة تمثل النص في لغة واحدة وإخراج وحدات بت تمثل النص نفسه بلغة أخرى. ويتمثل الهدف من البحث التلقائي بالشروح التوضيحية في إدخال أجزاء تمثل صورة وإخراج بتات تمثّل النص الذي يصفها (الشكل 5.2، اليمين).



الشكل 6.2: تأخذ البوابة المعروفة بـ NAND بتين اثنين، A و B كمدخلات وتحسب بتّاً واحداً هو C كمخرج. وفقاً للقاعدة التي تعتمد على C = 0 إذا A = B = 1 و C = 1، يمكن استخدام العديد من الأنظمة المادية كبوابات NAND. في المثال الأوسط، تُفسّر المُحوّلات على هيئة بتات بحيث 0 = مفتوحة، 1 = مغلقة، وعندما يُغلق كلا المحولين A و B، يفتح مغناطيس كهربائي المفتاح C. في المثال الموجود في أقصى اليمين، يُفسَّر الفولتية (الجهد الكهربائي) على أنها بت إذ 1 = 5 فولت، 0 = صفر فولتية، وعندما يكون السلكان A و B عند 5 فولت، يوصل الترانزستوران الكهرباءَ والسلك C ينخفض إلى صفر فولت تقريباً.

بعبارة أخرى، إذا كان بإمكانك تنفيذ دوال معقدة جدّاً، يمكنك بناء آلة ذكية قادرة على تحقيق أهداف بالغة التعقيد. هنا يقع سؤالنا عن كيف يمكن للمادة أن تكون ذكية في بؤرة تركيز أكثر وضوحاً: على وجه الخصوص، كيف يمكن لكتلة من مواد غبية أن تجري عمليات حوسبة لمهمة معقدة؟

فبدلاً من أن تظل ساكنة كحلقة ذهبية أو أي ذاكرة ثابت Static memorty أخرى، يجب أن تُظهِر **ديناميكية** معقدة بحيث تعتمد حالتها المستقبلية بطريقة معقدة (يمكن التحكم فيها/قابلة للبرمجة) على الحالة الراهنة، ويجب أن يكون ترتيب ذراتها أقل ترتيباً من مادة صلبة جامدة حيث لا يتغير شيء مثير للاهتمام، ولكن أكثر ترتيباً من سائل أو غاز. على وجه التحديد، نريد أن يكون للنظام الخاصية التي إذا وضعناها في حالة تُرمِّز Encode معلومات الإدخال، وتركناها تتطور وفقا لقوانين معينة لبعض الوقت، ومن ثم فسَّرنا الحالة النهائية الناتجة كمعلومات إخراج، فإن المُخرَج هو دالة الإدخال، وإذا كانت هذه هي الحال، فيمكننا القول إن النظام يجرى عمليات الحوسبة التي تُمليها دالتنا.

كمثال أول على هذه الفكرة، دعنا نستكشف الكيفية التي يمكننا بها بناء دالة بسيطة جدّاً (ولكنها مهمة أيضاً) تسمى بوابة NAND * من مادة بكماء عادية، تُدخل هذه الدالة اثنين من البتات ويكون المخرج بتا واحدا: فهي تُخرج 0 إذا كان كلا المدخلين1 ؛ وفي جميع الحالات الأخرى، فإن المُخرج هو 1. وإذا وصلنا مفتاحين على التوالي مع بطارية ومغناطيس كهربائي؛ فسيُفعَّل المغناطيس الكهربائي فقط إذا تَمَّ إغلاق المفتاح الأول والمفتاح الثاني («on»). لنضع مفتاحاً ثالثاً تحت المغناطيس الكهربائي، كما هو موضح في الشكل 6.2، بحيث يسحب المغناطيس المفتاح ويفتحه متى ما أُدير النظام، إذا فسرنا المبدأين الأولين كباقات الإدخال والبت الثالث كنقطة مخرج (مع 0 = مفتاح مفتوح، و 1 = مفتاح مغلق)، عندها تكون لدينا بوابة NAND: المفتاح الثالث مفتوح فقط إذا كان المفتاحان الأولان مغلقين. وهناك العديد من الطرق الأخرى الأكثر عملية لبناء بوابات NAND - مثلاً، باستخدام الترانزستورات كما هو موضح في الشكل 6.2، في أجهزة الحاسوب الحالية تُنشئ بوابات NAND عادة من الترانزستورات المجهرية والمكونات الأخرى التي يمكن حفرها تلقائيّاً في رقائق السيليكون.

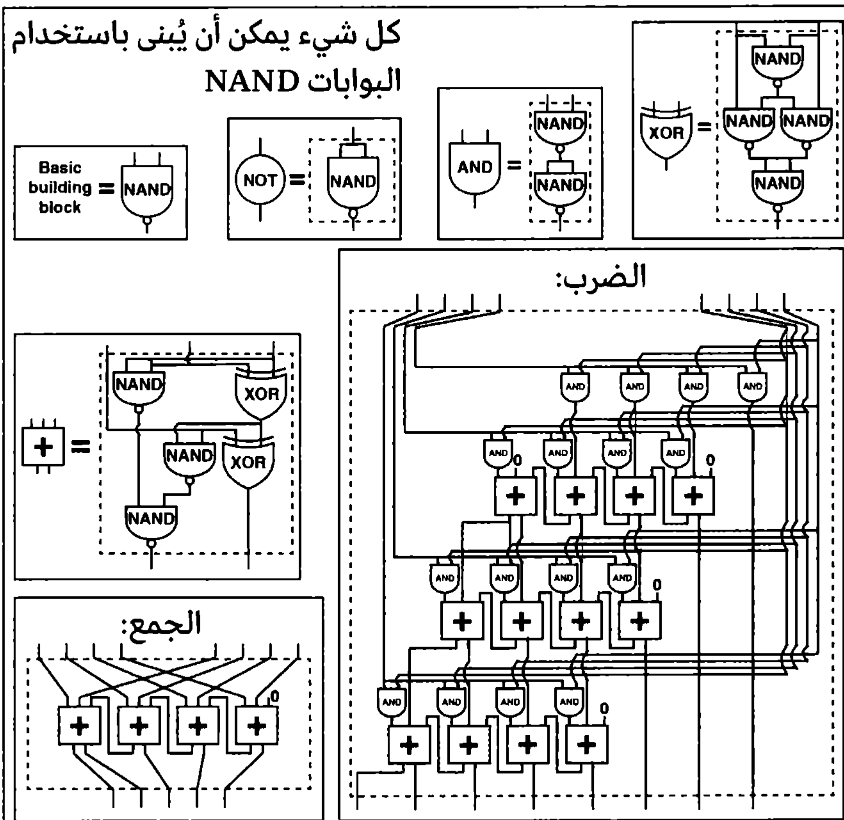
توجد نظرية رائعة في علوم الحاسوب تقول إن بوابات NAND عالمية، أي إنه يمكنك تنفيذ أي دالة محددة بشكل جيد بمجرد ربط البوابات NAND،** إذا كان بإمكانك بناء عدد كافٍ من بوابات NAND، يمكنك إنشاء جهاز يحسب أي شيء! في حالة رغبتك في التعرف على كيفية عمل ذلك، فقد أوضحت في الشكل 7.2 كيفية مضاعفة الأعداد باستخدام أي شيء باستثناء البوابات NAND.

صاغ باحثا المعهد MIT نورمان مارغولس Norman Margolus وتوماسو توفولي Tommaso Toffoli مصطلح كومبيوترونيوم ***Computronium*** لوصف أي مادة قد تؤدي حوسبات إحصائية، وقد رأينا للتو أن صنع الكومبيوترونيوم ليس بالضرورة أمراً صعباً: فالمادة تحتاج فقط إلى أن تكون قادرة على تنفيذ البوابات NAND المتصلة ببعضها البعض بأي طريقة مرغوب فيها، بالطبع، هناك أيضاً أنواع أخرى لا حصر لها من الكومبيوترونيوم. وأحد الأشكال البسيطة والعاملة أيضاً تنطوي على استبدال البوابات NAND بالبوابات NOR التي تُخرج 1 فقط عندما تكون المدخلات 0. وفي القسم التالي، سنستكشف الشبكات العصبية Neural networks، التي يمكنها أيضاً تنفيذ حوسبات عشوائية، أي تعمل ككومبيوترونيوم، أظهر العالِم ورائد الأعمال ستيفن ولفرام Stephen

* إن مصطلح NAND هو اختصار لـ NOT AND: فمخرج البوابة NAD هو 1 فقط إذا كان المدخل هو 1 والمدخل الثاني هو 1، وهكذا فإن مُخرَجات NAND هي عكس ذلك بالضبط.

** استخدم «دالة محددة تحديداً جيداً» Well- defined function لتعني ما يسميه علماء الرياضيات وعلماء الحاسوب «دالة قابلة للحوسبة» Computable function، أي دالة يمكن حوسبتها بواسطة حاسوب افتراضي يحتوي على ذاكرة ووقت غير محدودين، أثبت آلان تورينغ وتشيرش في برهانهما الشهير أن هناك أيضاً دوالّ يمكن وصفها ولكنها غير قابلة للحوسبة.

Wolfram أن الأمر نفسه ينطبق على الأجهزة البسيطة المسماة الأوتوماتا الخلوية Cellular automata التي تُحدِّث البتات بشكل مطّرد بناء على ما تفعله الخلايا الجيران، وفي عام 1936 أثبت رائد الحاسوب آلان تورينغ في ورقة بحث تاريخية أن آلة بسيطة (تعرف الآن بـ "آلة تورينغ العالمية" Universal Turing machine) يمكنها أن تتلاعب بالرموز على قطعة من شريط قد تنفذ أيضاً عمليات حوسبية عشوائية. باختصار، ليس فقط بإمكان المادة أن تنفِّذ عملية حوسبية مُعرَّفة تعريفاً جيداً، بل بإمكانها القيام بذلك بعدد كبير من الطرق المختلفة.



الشكل 7.2: يمكن إجراء أي حوسبة محددة تحديداً جيداً Well-defined computation من خلال الجمع بذكاء بين لا شيء سوى البوابات NAND، مثلاً، وحدات Modules الإضافة والضرب أعلاه كُلٌّ منها يُدخِل أعدادا ثنائية Binary numbers يتم تمثيلهما بأربعة بتات، ويُخرجان رقماً ثنائيّاً يُمثَّل بـ 5 و 8 بتات، على التوالي، الوحدات الصغيرة NOT و AND و XOR و + (التي تجمع ثلاث وحدات من البتات المنفصلة إلى رقم ثنائي يتألف من بتين اثنين) يتم إنتاجها بدورها من بوابات NAND. وفهم هذا الشكل التوضيحي فهماً تامّاً هو أمر بالغ الصعوبة وغير ضروري تماماً لمتابعة بقية هذا الكتاب. وقد أدرجت الشكل التوضيحي هنا فقط لتوضيح فكرة «العامة» - ولإرضاء هوسي الذاتي.

كما ذكرنا سابقاً، في تلك الورقة البحثية التي كتبها في عام 1936 أثبت تورينغ أيضاً أمراً أكثر عمقاً: إذا كان هناك نوع من الحاسوب يمكنه إجراء مجموعة معينة من العمليات، فإنه **حاسوب عام *Universal***: بمعنى أنه إذا تَمَّ توفير موارد كافية، فيمكنه أداء أي مهمة يؤديها أي حاسوب آخر. وقد أظهر أن جهاز تورينغ كان جهازاً عامّاً، ويرتبط بشكل وثيق بالفيزياء. وقد رأينا للتو أن هذه العائلة من أجهزة الحاسوب العامة تشتمل أيضاً على أشياء متنوعة مثل شبكة البوابات NAND وشبكة من الخلايا العصبية المترابطة. وفي الواقع، فقد جادل ستيفن وولفرام في أن معظم الأنظمة المادية غير التافهة، من نظم الطقس إلى العقول، ستكون أجهزة حاسوب عامة إذا كان بالإمكان بناؤها اعتباطيّاً لتكون أكبر وأطول مدى. وحقيقة أن الحوسبة نفسها يمكن إجراؤها بالضبط على أي جهاز حاسوب عام تعني أن الحوسبة **مستقلة عن المادة *Substrate-independent*** بالاستقلالية نفسها التي تكون عليها المعلومات: تكون لها حياة خاصة بها، بغض النظر عن مادتها الفعلية! لذا إذا كنت شخصية ذكية فائقة في لعبة حاسوب مستقبلية، فلن يكون لديك أي طريقة لمعرفة ما إذا كنت تلعب ضمن برنامج تشغيل ويندوز Windows على جهاز سطح مكتب أو نظام ماك Mac على حاسوب محمول أو هاتف آندرويد Android. لأنك ستكون مستقلاً عن المادة، ولن تكون لديك أيضاً أي طريقة لمعرفة نوع الترانزستورات التي كان المعالِج الدقيق Microprocessor يستخدمها.

لقد توصَّلْتُ إلى تقدير هذه الفكرة الحاسمة من استقلال المادة لأول مرة بفضل الأمثلة الجميلة في الفيزياء، الموجات، مثلاً: لديها خصائص مثل السرعة Speed والطول الموجي Wavelength والتردد Frequency. ونحن الفيزيائيين قادرون على دراسة المعادلات التي تطيعها الموجات دون الحاجة إلى معرفة ماهية المادة التي تتموج عبرها، وعندما تسمع شيئاً ما، فإنك ترصد موجات صوتية تسببها جزيئات ترتد في خليط الغازات الذي نسميه الهواء. ويمكننا حوسبة جميع أنواع الأشياء المثيرة للاهتمام حول هذه الموجات - كيف تتلاشى شدتها متناسبة مع مربع المسافة، أو كيف تنحني عند المرور عبر الأبواب المفتوحة، أو كيف ترتد من الجدران وتسبب الصدى - دون معرفة ممّا يتألف الهواء، في الواقع، نحن لا نحتاج حتى إلى معرفة أنه مصنوع من جزيئات: يمكننا تجاهل جميع تفاصيل الأكسجين والنيتروجين وثاني أكسيد الكربون، ... إلخ، لأن الخاصية الوحيدة لمادة الموجة هي المهمة، وتدخل في المعادلة الموجية الشهيرة هي عدد واحد يمكننا قياسه: سرعة الموجة، التي تبلغ في هذه الحالة نحو 300 متر في الثانية، بالطبع، فإن معادلة الموجة التي درستها لطلبتي في المعهد MIT في الفصل الدراسي بالربيع الماضي، كانت قد اكتُشِفَت واستُخدمت استخداماً موسعا حتى قبل أن يُثبِت الفيزيائيون أن الذرات والجزيئات موجودة!

يوضح مثال الموجة هذا ثلاث نقاط مهمة: أولاً، لا يعني استقلالية المادة أن المادة غير ضرورية، ولكن معظم تفاصيلها غير مهمة، ومن الواضح أنه لا يمكن أن

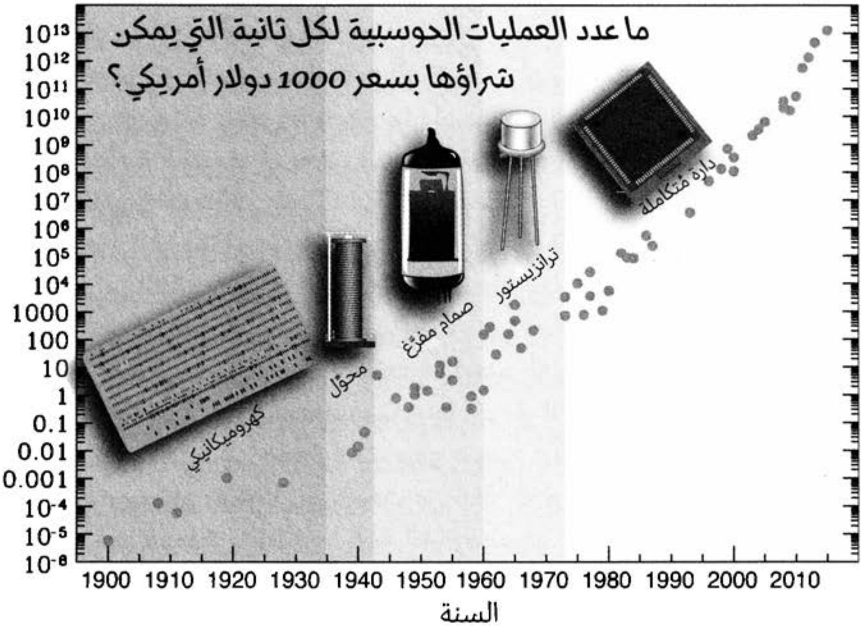
تكون لديك موجات صوتية في الغاز إذا لم يكن هناك غاز، ولكن أي غاز على الإطلاق سيفي بالغرض. وبالمثل، من الواضح أنه لا يمكن أن تكون لديك حوسبة من دون مادة، ولكن أي مادة ستفي بالغرض ما دام يمكن ترتيبها في البوابات NAND أو الخلايا العصبية المتصلة أو بعض لبنات البناء الأخرى التي تُمكّن الحوسبة العامة؛ ثانياً، إن ظاهرة الاستقلال عن المادة تأخذ منحنى خاصّاً بها، مستقلة عن مادتها، ويمكن للموجة أن تنتقل عبر بحيرة، على الرغم من أن أيّاً من جزيئات الماء لا تفعل ذلك، فهي في الغالب تتحرك إلى الأعلى والأسفل، مثلما يُشكّل المشجعون حركة «الموجة» في الملعب الرياضي؛ ثالثاً، غالباً ما يكون الجانب المستقل الذي نرغب فيه فقط هو الجانب المستقل عن المادة: فعادة ما يهتم راكب الموجات (الركمجة) بموقع الموجة وارتفاعها أكثر من اهتمامه بتكوينها الجزيئي التفصيلي، فقد رأينا كيف كان هذا صحيحا بالنسبة إلى المعلومات، وهو صحيح أيضاً بالنسبة إلى الحوسبة: إذا كان هناك مبرمجان يشتركان في البحث عن خطأ البرمجة Bug في كودهما Code؛ فمن المحتمل أنهما لا يناقشان تركيبة الترانزستورات.

لقد توصلنا الآن إلى إجابة لسؤالنا الافتتاحي حول كيف يمكن للأشياء المادية الملموسة أن تثير شيئاً يُشْعَر بأنه غير ملموس ومُجرَّد وأثيري مثل الذكاء: إنه شعور غير مادي لأنه مستقل عن المادة، ويجري في منحنٍ خاص به لا يعتمد على التفاصيل المادية وليس هو بانعكاس لها. باختصار، الحوسبة هي نمط Pattern في ترتيب الزمكان للجسيمات، وليست الجسيمات ولكن النمط هو الذي يُهمُّ حقّاً! المادة لا تهم.

وبعبارة أخرى، فإن الأجهزة (العتاد) Hardware هي المادة والبرمجيات Software هي النمط. إن استقلالية الحوسبة هذه تعني ضمناً أن الذكاء الاصطناعي AI ممكن: الذكاء لا يتطلب ذرات من اللحم أو الدم أو الكربون.

وبسبب استقلالية المادة هذه، تمكَّن المهندسون الماهرون من الاستعاضة عن التقنيات الموجودة داخل أجهزة الحاسوب الخاصة بنا بشكل مطّرد بأجهزة أفضل بشكل كبير، دون تغيير البرمجيات. وقد كانت النتائج مذهلة تماماً مثلها مثل تطورات شرائح الذاكرة. فكما هو موضح في الشكل 8.2، فإن تكلفة الحوسبة لا تزال تنخفض إلى النصف تقريباً كل عامين. وقد استمر هذا الاتجاه منذ أكثر من قرن؛ مما أدى إلى خفض تكلفة الحاسوب إلى مليون مليون مليون مرة (10^{18}) مرة منذ ولادة جدتي. وإذا رخص كل شيء إلى مليون مليون مليون ضعف، فإن واحداً في المئة من السنت Cent سيمكّنك من شراء جميع السلع والخدمات المنتجة على الأرض هذا العام. وهذا الانخفاض المفاجئ في التكاليف هو، بالطبع، سبب رئيسي لشيوع الحوسبة في كل مكان في هذه الأيام، بعد أن انتشرت خارجة من مرافق الحوسبة ذات المباني الكبيرة إلى منازلنا وسياراتنا وجيوبنا، بل حتى في أماكن غير متوقعة مثل الأحذية الرياضية.

لماذا تستمر تقنيتنا في مضاعفة قوتها على فترات منتظمة، وهو ما يصف ما يسميه علماء الرياضيات النمو المتسارع Exponential growth؟



الشكل 8.2: منذ عام 1900 صارت الحوسبة أرخص بنصف التكلفة كل عامين تقريباً. ويُظهر الرسم البياني قدرة الحوسبة المقاسة بحسابات النقطة العائمة Floating point operation في الثانية (اختصاراً: الفلوبات FLOPS) التي يمكن شراؤها بمبلغ 1000 دولار أمريكي.[3] والحوسبة المُعيَّنة التي تُحدد حساب النقطة العائمة تعادل نحو 10^5 عمليات منطقية أساسية Elementary logical operations مثل تقلبات البتات أو تقييمات NAND.

بالطبع، لماذا يحدث ذلك ليس فقط من حيث تصغير الترانزستور (وهو ما يعرف بقانون مور Moore's law)، ولكن أيضاً بشكل عام في الحوسبة ككل (الشكل 8.2)، وفي الذاكرة (الشكل 4.2) وفي عدد كبير من التقنيات الأخرى من سَلْسَلَة الجينوم إلى تصوير الدماغ؟ يُطلق راي كورزويل Ray Kurzweil على ظاهرة المضاعفة المستمرة هذه «قانون العوائد المتسارعة Law of accelerating returns».

وجميع الأمثلة على المضاعفة المستمرة التي أعرفها في الطبيعة لها السبب الأساسي نفسه، وهذه المضاعفة المستمرة التقنية ليست استثناء: كل خطوة تخلق الخطوة التالية، مثلاً، خضعت نفسك لنمو متسارع بعد الحمل بك: فكل خلية من خلاياك انقسمت وأدت إلى إنشاء خليتين يوميّاً تقريباً؛ مما أدى إلى زيادة العدد الإجمالي للخلايا يوماً بعد يوم مثل 1 و 2 و 4 و 8 و 16 وهكذا. ووفقاً للنظرية العلمية الأكثر شيوعاً لأصولنا الكونية، المعروفة **بالتضخم *Inflation***، فإن كوننا الطفل نما ذات يومٍ أضعافاً مضاعفة مثلما فعلت أنت، فضاعف حجمه مراراً وتكراراً على فترات منتظمة حتى إن شذرة أصغر بكثير وأخف من الذرة نمت إلى أكثر من جميع المجرات التي رأيناها على الإطلاق بتلسكوباتنا. مجدداً، كان السبب هو العملية التي تؤدي فيها كل خطوة مضاعفة إلى الخطوة التالية. وهذه هي

الطريقة التي تتقدم بها التكنولوجيا أيضاً: فبمجرد أن تتضاعف قوة التكنولوجيا، يمكن استخدامها في كثير من الأحيان لتصميم وبناء تكنولوجيا بدورها ذات قوة مضاعفة؛ مما يؤدي إلى مضاعفة القدرة التي تتضاعف وفقاً لروح قانون مور.

والأمر الآخر الذي يحدث بمثل انتظام مضاعفة قوتنا التكنولوجية هو ظهور ادعاءات مفادها أن المضاعفة شارفت على النهاية. نعم، سيصل قانون مور بالطبع إلى نهاية، وهذا يعني أن هناك حدّاً ماديّاً للمدى الذي يمكن تصغير الترانزستورات إليه. ولكن البعض يفترضون خطأً أن قانون مور مرادف للمضاعفة المستمرة لقوتنا التكنولوجية. فعلى العكس من ذلك، يشير راي كورزويل إلى أن قانون مور لا ينطوي على النموذج Paradigm التكنولوجي الأول بل الخامس لتحقيق النمو المتسارع في الحوسبة. وكما هو موضح في الشكل 8.2: كلما توقفت إحدى التقنيات عن التطور، استعضنا عنها بتقنية جديدة أفضل، وعندما لم يعد بإمكاننا الاستمرار في تقليص الصمامات المفرغة Vacuum tubes، استعضنا عنها بالترانزستورات Transistors ثم الدارات المتكاملة Integrated circuits، حيث تتحرك الإلكترونات في بعدين. وعندما تصل هذه التكنولوجيا إلى حدودها، ستكون هناك العديد من البدائل الأخرى التي يمكننا تجربتها - مثلاً، استخدام دارات ثلاثية الأبعاد Three-dimensional circuits واستخدام شيء آخر غير الإلكترونات للقيام بمهامنا.

لا أحد يعرف على وجه اليقين ما ستكون عليه المادة الحوسبية الرائجة المقبلة، ولكننا نعرف أننا لسنا بعد بالقرب من الحدود التي تفرضها قوانين الفيزياء، فقد توصل زميلنا في المعهد MIT، سيث لويد Seth Lloyd، إلى ما هو هذا الحد الأساسي. وكما سنستكشف بتفصيل أكبر في الفصل السادس، فإن هذا الحد هو عدد ضخم يصل إلى 33 ضعفاً (10^{33} مرة) أحدث تقنيات ما تستطيع قطعة من المادة فعله من عمليات حوسبية، لذا حتى إذا استمرت بنا الحال بمضاعفة قوة أجهزة الحاسوب لدينا كل عامين؛ فسوف يستغرق الأمر أكثر من قرنين حتى نصل إلى هذه الحدود النهائية.

وعلى الرغم من أن جميع أجهزة الحاسوب العامة قادرة على إجراء الحوسبات نفسها، فإن بعضها أكثر كفاءة من غيرها. مثلاً، لا تتطلب حوسبة ملايين عمليات الضرب الملايين من وحدات الضرب المنفصلة التي تتألف من ترانزستورات منفصلة كما في الشكل 6.2: بل تحتاج إلى وحدة واحدة فقط، إذ يمكنها استخدامها عدة مرات متتالية باستخدام المدخلات المناسبة. وبمثل روح الكفاءة هذه، تستخدم معظم الحواسيب الحديثة نموذجاً يتم فيه تقسيم الحوسبات إلى خطوات زمنية متعددة، يتم خلالها خلط المعلومات ذهاباً وإياباً بين وحدات الذاكرة ووحدات الحوسبة. وطوّرت هذه البنية الحوسبية بين عامي 1935 و 1945 من قبل رواد الحاسوب، بمن في ذلك آلان تورينغ وكونراد زوس Konrad Zuse وبرسبر إيكيرت Presper Eckert وجون موتشلي John Mauchly وجون فون نويمان John von Neumann، وبشكل أكثر تحديداً، تخزن ذاكرة الحاسوب كلا من البيانات والبرمجيات (برنامج: أي قائمة إرشادات لما يجب فعله بالبيانات). وفي كل خطوة زمنية تنفذ وحدة المعالجة المركزية (CPU) التعليمات التالية في البرنامج، التي تحدد بعض الدوال

البسيطة لتطبيقها على جزء ما من البيانات. وجزء الحاسوب الذي يتتبع ما يجب فعله بعد ذلك هو مجرد جزء آخر من ذاكرته، يسمى عداد البرنامج Program counter، الذي يخزن رقم السطر الحالي في البرنامج. وللانتقال إلى الأمر التالي، ببساطة أضفْ سطراً آخر إلى عداد البرنامج، للقفز إلى سطر آخر من البرنامج. ببساطة انسخ هذا العدد إلى عداد البرنامج - وهذا هو الأسلوب الذي تُنفَّذ به ما يعرف بالعبارات الشرطية «if» (إذا) والحلقات Loop.

غالباً ما تكتسب أجهزة الحاسوب اليوم سرعة إضافية عن طريق **المعالجة المتوازية** ***Parallel processing***، التي تُقلِّل بذكاء بعض إعادة استخدام الوحدات: إذا كان من الممكن تقسيم الحوسبة إلى أجزاء يمكن تنفيذها بالتوازي (لأن مدخل جزء واحد لا يتطلب مخرجاً من آخر)، يمكن حينها حوسبة الأجزاء في وقت واحد من قبل أجزاء مختلفة من الأجهزة.

أما الحاسوب المتوازي الأقصى؛ فهو **حاسوب كمّيّ** ***Quantum computer***، ورائد الحوسبة الكمّيّة دافيد دويتش David Deutsch يثير الجدل بقوله إن «الحواسيب الكمّيّة تتشارك في المعلومات مع أعداد هائلة من النسخ من نفسها عبر الكون المتعدد Multiverse»، ويمكنها أن تحصل على إجابات أسرع هنا في كوزموسنا من خلال الحصول على المساعدة من هذه النسخ الأخرى،[4] لا نعرف بعد ما إذا كان من الممكن بناء حاسوب كمّيّ قادر على المنافسة تجاريّاً خلال العقود القادمة؛ لأنه يعتمد على ما إذا كانت الفيزياء الكمّيّة تعمل كما نعتقد، وعلى قدرتنا على التغلب على التحديات التقنية الهائلة. ولكن الشركات والحكومات في جميع أنحاء العالم تراهن بعشرات الملايين من الدولارات سنويّاً على الاحتمال. وعلى الرغم من أن الحواسيب الكمّيّة لا يمكنها تسريع العمليات الحوسبية الاعتيادية، فإنه قد تَمَّ تطوير خوارزميات Algorithms ذكية قد تسرع بشكل كبير من أنواع مُعيَّنة من الحَوْسبات، مثل فك أنظمة التشفير (التعمية) Cryptosystems وتدريب شبكات التدريب. كما يمكن لحاسوب كمّيّ أن يحاكي بفعالية سلوك الأنظمة الميكانيكية الكمّيّة، بما في ذلك الذرات والجزيئات والمواد الجديدة، لتحل محل القياسات في مختبرات الكيمياء بالطريقة نفسها التي استبدلت بها التقليدية القياسات في أنفاق الرياح Wind tunnels بالمحاكاة على أجهزة الحاسوب.

ما هو التعلم؟

على الرغم من أن آلة حاسبة الجيب يمكن أن تتغلب عليّ في مسابقة حساب Arithmetic contest، فإنها لن تحسّن سرعتها أو دقتها مهما مارست الحساب، إنها لا تتعلم: مثلاً، في كل مرة أضغط فيها على زر الجذر التربيعي عليها، فإنها تحسب بالضبط الدالة نفسها بالطريقة نفسها بالضبط. وبالمثل، فإن أول برنامج حاسوب غلبني في الشطرنج لم يتعلم قطُّ من أخطائه، بل نفذ فقط دالة صممها مبرمجه الذكي لحوسبة حركة تالية جيدة. وفي المقابل، عندما خسر ماغنوس كارلسن Magnus Carlsen لعبته الأولى للشطرنج في سن الخامسة، شرع في عملية تَعلُّم جعلت منه بطل العالم في الشطرنج بعد 18 عاماً.

وقد يُجادَل في أن القدرة على التعلم هي الجانب الأكثر سحراً من الذكاء العام، فقد رأينا بالفعل الكيفية التي يمكن بها أن تتذكر كتلة من المادة التي قد تبدو غبية وتجري عمليات حوسبية، ولكن كيف لها أن تتعلم؟ وقد رأينا أن العثور على إجابة لسؤال صعب يُوازي حوسبة دالة، وأن المادة المرتبة بشكل مناسب يمكن أن تحسب Calculate أي دالة قابلة للحوسبة. وعندما صنع البشر أولاً حاسبات الجيب وبرمجيات الشطرنج، قمنا نحن بعملية الترتيب. ولكي تتعلم المادة، يجب عليها بدلاً من ذلك أن تعيد ترتيب نفسها لتصبح أفضل وأفضل في حوسبة الدالة المطلوبة - ببساطة عن طريق إطاعة قوانين الفيزياء.

لإزالة الغموض المحيط بعملية التعلم، لننظر أولاً في كيفية معرفة نظامٍ مادي بسيط جدّاً للأرقام المُشكِّلة للعدد π وغيره من الأعداد أخرى. رأينا آنفاً الكيفية التي يمكن بها استخدام سطح يحوي العديد من الوديان كجهاز ذاكرة (انظر: الشكل 3.2) : مثلاً، إذا كان قعر أحد الوديان يقع في موضع يعادل $x = \pi \approx 3.14159$ وليس هناك أي أودية أخرى قريبة، يمكنك وضع الكرة عند $x = 3$ ومشاهدة النظام يُجري عمليات حوسبة الكسور العشرية المفقودة عن طريق ترك الكرة تتدحرج إلى الأسفل. الآن، لنفترض أن السطح مصنوع من الطين الطري ويبدأ من حالة يكون فيها مسطحاً تماماً كلوح أملس، وإذا قام بعض محبي الرياضيات بترتيب الكرة مراراً وتكراراً في مواقع تتألف من أعدادهم المفضلة؛ فستعمل الجاذبية تدريجيّاً على حفر الوديان في هذه المواقع، وبعد ذلك يمكن استخدام سطح الطين لتذكر هذه الذكريات المُخزَّنة. وبعبارة أخرى، فقد **تعلم** سطح الطين حوسبة أرقام الأعداد من مثل π.

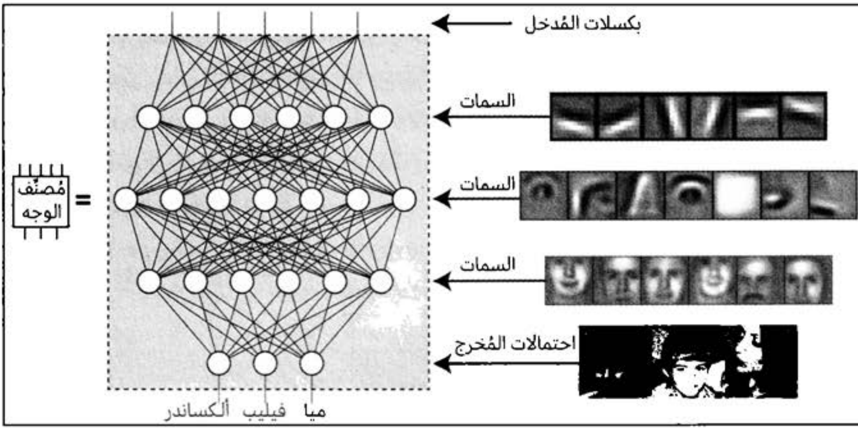
أما النظم الفيزيائية الأخرى، مثل الأدمغة؛ فيمكنها أن تتعلم بشكل أكثر كفاءة بناء على الفكرة نفسها. وأظهر جون هوبفيلد John Hopfield أن شبكته من الخلايا العصبية المتشابكة المذكورة آنفاً قد تتعلم بطريقة مماثلة: إذا وضعتها مراراً في حالات معينة؛ فستتعلم هذه الحالات تدريجيّاً وتعود إليها من أي حالة مجاورة. إذا كنت قد رأيت كل فرد من أفراد عائلتك عدة مرات؛ فإن ذكريات ما يبدون عليه يمكن إثارتها بأي شيء متعلق بهم.

لقد غيَّرت الشبكاتُ العصبية كلّاً من الذكاءَ البيولوجي والاصطناعي، وبدأت في الآونة الأخيرة بالسيطرة على الحقل الفرعي للذكاء الاصطناعي المعروف **بتَعلُّم الآلة** ***Machine learning*** (دراسة الخوارزميات التي تتحسن من خلال التجربة). وقبل التعمق أكثر في كيفية تعلّم هذه الشبكات، دعنا نفهم أولاً كيفية قيامها بالحوسبة. والشبكة العصبية هي ببساطة مجموعة من العصبونات (الخلايا العصبية) Neurons المترابطة التي يمكنها التأثير في سلوك بعضها البعض، ويحتوي دماغك على عدد من الخلايا العصبية يعادل تقريباً عدد النجوم في مجرتنا: في حدود مئة بليون. في المتوسط، كل خلية من هذه العصبونات متصلة بنحو ألف خلية أخرى عبر وصلات تسمى مشتبكات عصبية Synapses. وقوة هذه الوصلات التي يصل عددها تقريباً إلى مئة تريليون مشتبك عصبي هي التي تُرمِّز Encode معظم المعلومات في دماغك.

ويمكننا رسم شبكة عصبية كمجموعة من النقاط التي تمثل العصبونات الموصولة بخطوط تمثل المشتبكات العصبية (انظر: الشكل 9.2)، العصبونات في العالم الحقيقي هي أجهزة كهروكيميائية معقدة جدّاً لا تشبه شيئاً مما في هذا الرسم التخطيطي: فهي تتألف من أجزاء مختلفة بأسماء مثل المحاور Axons والزوائد الشجيرية Dendrites. وهناك العديد من الأنواع المختلفة من العصبونات التي تعمل بمجموعة متنوعة من الطرق. والتفاصيل الدقيقة لكيفية تأثير النشاط الكهربائي في عصبون وزمنه في العصبونات الأخرى لا يزال موضوع دراسة حثيثة. ولكن، أظهر باحثو الذكاء الاصطناعي أن الشبكات العصبية لا يزال بإمكانها أن تصل إلى مستوى أداء الإنسان في العديد من المهام المعقدة بشكل ملحوظ حتى ولو تجاهل المرء جميع هذه التعقيدات وتمت الاستعاضة عن العصبونات البيولوجية الحقيقية بعصبونات اصطناعية بسيطة جدّاً ومتطابقة وتتبع قواعد بسيطة جدّاً، حاليا النموذج الأكثر شيوعا **لشبكة عصبية اصطناعية *Artificial neural network*** يُمثِّل حالة كل عصبون بعدد واحد وقوة كل مشتبك بعدد واحد. وفي هذا النموذج Model يُحدِّث كل عصبون حالته في فترات زمنية منتظمة ببساطة بحساب متوسط المدخلات من جميع العصبونات المتصلة، وتقييمها وفقاً لقوة المشتبك العصبي، وإضافة ثابت Constant اختياريّاً، ثم تطبيق ما يسمى **دالة التنشيط *Activation function*** على النتيجة لحوسبة الحالة التالية.* أسهل طريقة لاستخدام الشبكة العصبية كدالة من خلال **التغذية الأمامية *Feedforward***، مع تدفق المعلومات في اتجاه واحد فقط، كما في الشكل 9.2، إضافة المدخل إلى الدالة إلى طبقة من العصبونات في الأعلى واستخراج المُخْرَج من طبقة من العصبونات في القاع.

ونجاح هذه الشبكات العصبية الاصطناعية البسيطة هو مثال آخر على الاستقلال عن المادة: فالشبكات العصبية لديها قدرة حوسبية كبيرة تبدو مستقلة عن التفاصيل التافهة لمبناها. وفي الواقع، فإن جورج سيبنك George Cybenko، وكورت هورنيك Kurt Hornik، وماكسويل ستينكومب Maxwell Stinchcombe، وهالبرت وايت Halbert White أثبتوا أمراً رائعاً في عام 1989: هذه الشبكات العصبية البسيطة هي عامة Universal، بمعنى أنها تستطيع أن تحسب أي دالة بدقة اعتباطيّاً، ببساطة عن طريق ضبط أعداد شدة المشتبكات العصبية وفقاً لذلك. وبعبارة أخرى، فإن التطور ربما لم يجعل خلايانا العصبية البيولوجية معقدة جدّاً لأن التعقيد ضروري، بل لأنه أكثر كفاءة - ولأن التطور، مقارنة بالمهندسين البشريين، لا يكافِئ تصاميم بسيطة وسهلة الفهم.

* إذا كنت تحب الرياضيات، هناك خياران شائعان لدالة التنشيط Activation function هذه هما: ما تعرف بالدالة السينيَّة Sigmoid function حيث $\sigma(x) \equiv 1/(1 + e-x)$ ودالة المنحدر Ramp function حيث $\sigma(x)=\max\{0, x\}$. على الرغم من أنه قد بُرهِن على أن أي دالة تكفي ما دامت ليست خطية Not linear (خط مستقيم)، فإن نموذج هوبفيلد Hopfield' model الشهير $\sigma(x)= -1$ إذا كان $x < 0$ و$\sigma(x)= 1$ إذا كان $x \geq 0$، إذا تم تخزين حالات العصبونات في مُتّجِه Vector، عندما تُحدَّث الشبكة ببساطة بضرب هذا المُتّجِه في مصفوفة تخزين المشتبكات العصبية ثم تطبيق الدالة σ على جميع العناصر.



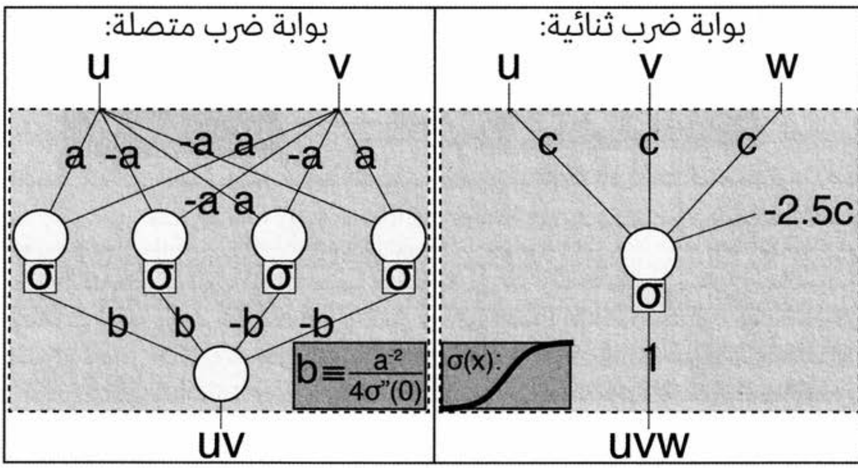
الشكل 9.2: يمكن لشبكة من الخلايا العصبية (العصبونات) حساب الدالات تماماً مثلما تستطيع شبكة من البوابات NAND. مثلاً، تُدرَّب شبكات عصبية اصطناعية على أعداد مُدخلات تمثل سطوع وحدات بكسل مختلفة في الصور وأعداد مخرجات تمثل احتمال أن تُظهر الصورة أشخاصاً مختلفين. هنا يحسب كل عصبون اصطناعي (الدائرة) المجموع الموزون Weighted sum للأعداد المرسلة عبر الوصلات (خطوط من الأعلى، ويطبق دالة بسيطة ويمرر المعلومات إلى الأسفل، وكل طبقة لاحقة تحسب سمات المستوى الأعلى. تحتوي شبكات التعرف على الوجوه النموذجية على مئات الآلاف من العصبونات، ويعرض الشكل مجرد حفنة للتوضيح.

وعندما أدركت هذا لأول مرة، كنت في حيرة من الكيفية التي قد يحسب بها شيء بسيط جدّاً شيئاً معقداً بشكل اعتباطي. مثلاً، كيف يمكنك حوسبة شيء بسيط مثل الضرب، عندما يكون كل ما يُسمح لك به هو حوسبة المجموع الموزون Weighted sum وتطبيق دالة ثابتة واحدة؟ إذا كنت تريد تذوق كيفية عمل هذا، يوضح الشكل 10.2 الكيفية التي يمكن بها لمجرد خمس عصبونات فقط أن تضرب رقمين عشوائيين أحدهما بالآخر، والكيفية التي يمكن بها لعصبون واحدة أن يضرب ثلاثة بتات ببعضها البعض.

وعلى الرغم من أنه يمكنك إثبات إمكانك حوسبة أي شيء نظريّاً باستخدام شبكة عصبية كبيرة بشكل اعتباطي، فإن الدليل لا يقول أي شيء حول ما إذا كان يمكنك فعل ذلك من الناحية العملية باستخدام شبكة ذات حجم معقول. وفي الواقع، كلما فكرت أكثر في الأمر، شعرت بالحيرة أكثر من أن الشبكات العصبية تؤدي عملها بشكل جيد.

مثلاً، لنفترض أننا نرغب في تصنيف صور بالدرجات الرمادية والمؤلفة بحجم ميغل بيكسل Megapixel إلى فئتين، مثل القطط أو الكلاب. إذا استطاع كل واحد من مليون بكسل أن يأخذ واحداً من، مثلاً، 256 قيمة، فهناك $256^{1000000}$ صورة محتملة، ولكل واحدة منها، نرغب في حوسبة احتمال أن تكون صورة قط. وهذا يعني أن الدالة الاعتباطية التي تُدخِل صورة وتُخرج احتمالات مُحدَّدة بقائمة من الاحتمالات $256^{1000000}$، أي، عدد أكبر من أعداد الذرات الموجودة في الكون (نحو 10^{78}). ولكن، إن الشبكات العصبية التي لديها

بالكاد آلاف أو ملايين المقاييس تتمكن بطريقة ما من أداء مهام التصنيف هذه بشكل جيد. كيف يمكن للشبكات العصبية الناجحة أن تكون «رخيصة»، بمعنى أنها تتطلب القليل من المقاييس؟ بعد كل شيء، يمكنك إثبات أن شبكة عصبية صغيرة بما يكفي لتكون داخل الكون ستفشل في تقريب جميع الدوال تقريباً، وتنجح فقط في جزء ضئيل جدّاً من جميع المهام الحوسبية المعهودة إليها.



الشكل 10.2: كيف يمكن للمادة أن تقوم بالضرب، ولكن ليس باستخدام البوابات NAND كما في الشكل 7.2 بل العصبونات. والنقطة الأساسية هي أنها لا تتطلب التفاصيل التالية، وأن العصبونات لا تقوم بالحسابات (الاصطناعية والبيولوجية) فحسب، بل إن الضرب يتطلب عدداً أقل من العصبونات مقارنة بالبوابات NAND. **وتم تفاصيل اختيارية للمهووسين بالرياضيات:** الدوائر تقوم بعمليات الجمع، والمربعات تطبق الدالة σ، والخطوط تُضرب في الثوابت المكتوبة عليها. فالمدخلات هي أعداد حقيقية (اليسار) وبتات (اليمين)، وتغدو عمليات الضرب صحيحة عشوائية مع $a \to 0$ (اليسار) و $c \to \infty$ (اليمين). والشبكة إلى اليسار صالحة لأي دالة σ (x) المنحنية عند المنشأ (مع المشتقة الثانية ($\sigma''(0)\neq 0$) التي يمكن البرهنة عليها باستخدام توسعة تايلور σx (x) (Taylor expanding)، الشبكة الصحيحة ستتطلب أن الدالة σ (x) تقارب 0 و1 عندما x تصبح صغيرة جدّاً وكبيرة جدّاً، على الترتيب، والذي يتضح بالإشارة إلى أن $uvw = 1$ فقط إذا $u+v+w=3$. (هذه الأمثلة هي من ورقة كتبتُها مع طالبَي هنري لين Henry Lin وديفيد رولنيك David Rolnick، «لماذا يعمل التعلم العميق والرخيص بهذه الجودة؟» ?Why Does Deep and Cheep Learning Work So Well، ويمكن الاطلاع عليها على http://arxiv.org/abs/1608.08225). وبالجمع بين العديد من عمليات الضرب (كما في الأعلى) وعمليات الجمع، يمكنك حساب أي معادلة متعددة الحدود Polynomial المعروفة أنها قادرة على مقاربة أي دالة سلسة.

استمتعت كثيراً وأنا أتفكر في هذا وفي الأسرار المتصلة بها مع طالبي لي هنري لين. أحد الأمور التي أنا في أشد الامتنان لها في الحياة هي فرصة التعاون مع طلبة مدهشين، وهنري واحد منهم. وعندما دخل إلى مكتبي أول مرة ليسأل عمّا إذا كنت مهتماً بالعمل معه،

فكرت في نفسي أنه من الأنسب لي أن أسأل عما إذا كان مهتماً بالعمل معي: هذا الطفل المتواضع الودود والمبهج من شريفيبورت، لويزيانا، كان قد كتب ثماني أوراق علمية، وفاز بجائزة فوربس لمن تحت الثلاثين Forbes 30- Under-30، وأعطى محادثة TED حصلت على أكثر من مليون مشاهدة - عندما كان عمره 20 عاماً فقط! بعد عام، كتبنا ورقة توصلت إلى استنتاج مثير للدهشة: السؤال عن سبب عمل الشبكات العصبية جيداً لا يمكن الإجابة عنه بالرياضيات فقط؛ لأن جزءاً من الإجابة يكمن في الفيزياء. فقد وجدنا أن صنف الدوال التي تقذفنا بها قوانين الفيزياء وتجعلنا مهتمين بالحوسبة هي أيضاً من صنف صغير إلى درجة مدهشة؛ لأن قوانين الفيزياء هي أيضاً بسيطة إلى درجة مدهشة لأسباب لا تزال غير مفهومة تماماً. إضافة إلى ذلك، فإن الجزء الصغير من الدوال التي يمكن للشبكات العصبية حوسبتها يشبه إلى حد كبير الجزء الصغير الذي تجعلنا الفيزياءُ مهتمين به! وقد قمنا أيضاً بتوسيع نطاق العمل السابق الذي يوضح أن شبكات التعلم العميق العصبية (تسمى «عميقة» Deep إذا كانت تحتوي على العديد من الطبقات) هي أكثر كفاءة بكثير من الشبكات الضحلة للعديد من هذه الدوال موضع الاهتمام. مثلاً، إلى جانب طالب آخر رائع في المعهد MIT، ديفيد رولينيك، أظهرنا أن المهمة البسيطة المتمثلة في مضاعفة أعداد n تتطلب وجود أعداد هائلة 2n من العصبونات لشبكة من طبقة واحدة فقط، لكنها لا تتطلب سوى 4n عصبونات في شبكة عميقة. وهذا يساعد على تفسير ليس فقط لماذا الشبكات العصبية هي الآن البدعة الطاغية بين باحثي الذكاء الاصطناعي، ولكن أيضاً لماذا تطورت الشبكات العصبية في أدمغتنا: إذا طورنا العقول للتنبؤ بالمستقبل، فمن المنطقي أن نطور بنية حوسبية جيدة على وجه التحديد في حوسبة تلك المشكلات التي تهم العالم المادي.

الآن بعد أن استكشفنا كيفية عمل الشبكات العصبية وحوسبتها، دعنا نَعُد إلى السؤال عن الكيفية التي تَتَعَلَّم بها، على وجه التحديد، كيف يمكن للشبكة العصبية أن تتحسن في الحوسبة عن طريق تحديث مشتبكاتها العصبية؟

وفي كتابه المؤثر فيما تلاه والصادر في عام 1949، **تنظيم السلوك: نظرية عصبية نفسية** ***The Organization of Behavior: A Neuropsychological Theory***، جادل العالم النفسي الكندي دونالد هيب Donald Hebb في أنه إذا كان هناك عصبونان أحدهما قريب من الآخر ينشطان في كثير من الأحيان («إطلاق الإشارة العصبية» Firing) في الوقت نفسه، فإن اقترانهما المشتبكي العصبي سيصبح أقوى، بحيث يتعلمان المساعدة لتحفيز أحدهما الآخر - فكرة عبّر عنها شعار شائع Fire together, wire together (تُطلِق معاً، موصولة معاً)، على الرغم من أن التفاصيل حول كيفية تعلم الأدمغة الفعلية لا تزال بعيدة عن الفهم. وقد أظهرت الأبحاث أن الإجابات في كثير من الحالات أكثر تعقيداً، غير أنه أُثبِت أيضاً أنه حتى هذه القاعدة التعليمية البسيطة (المعروفة بالتعلّم الهيبياني Hebbian learning) تسمح للشبكات العصبية بتعلم أشياء مثيرة للاهتمام. وأظهر جون هوبفيلد أن التعلّم الهيبياني سمح لشبكته العصبية الاصطناعية المبسطة بتخزين الكثير من الذكريات

المعقدة ببساطة عن طريق التعرض لها مراراً وتكراراً. عادة ما يُعرف مثل هذا التعرض للمعلومات Information التي يمكن التعلُّم منها «بالتدريب» Training عند الإشارة إلى الشبكات العصبية الاصطناعية (أو إلى الحيوانات أو الأشخاص الذين يُعلَّمون مهارات ما)، على الرغم من أن ألفاظاً مثل «الدراسة» Studying أو «التعليم» Education أو «الخبرة» Experience قد تكون ملائمة أيضاً لوصف ذلك. وتميل الشبكات العصبية الاصطناعية التي تدعم أنظمة الذكاء الاصطناعي AI systems الحالية إلى الاستعاضة عن التعلّم الهيبياني بقواعد تعلم أكثر تعقيداً بأسماء مهوسة مثل «الانتشار الراجع» Backpropagation و«الانحدار العشوائي للتدرج» Stochastic gradient descent. ولكن الفكرة الأساسية هي نفسها: هناك قاعدة بسيطة حتمية، مثل قانون الفيزياء، يتم من خلالها تحديث حالة المشتبكات العصبية مع مرور الوقت، كما لو كان ذلك عن طريق السحر. إن هذه القاعدة البسيطة قد تجعل الشبكة العصبية تتعلم الحوسبات المعقدة بشكل ملحوظ إذا استخدم التدريب كميات كبيرة من البيانات Data، ولا نعرف حتى الآن بدقة قواعد التعلم التي تستخدمها أدمغتنا، ولكن مهما كانت الإجابة، فليس هناك ما يشير إلى أنها تنتهك قوانين الفيزياء.

ومثلما تكتسب معظم أجهزة الحاسوب الرقمية Dgital computers الكفاءة عن طريق تقسيم عملها إلى خطوات متعددة وإعادة استخدام الوحدات الحوسبية Computational modules عدة مرات، فإن الكثير من الشبكات العصبية والبيولوجية الاصطناعية تقوم بذلك أيضاً. وتحتوي الأدمغة على أجزاء هي ما يُطلق عليه علماء الحاسوب **بالشبكات العصبية المتكررة *Recuurent neural network*** وليس شبكات التغذية الأمامية Feedforward neural networks، حيث يمكن أن تتدفق المعلومات في اتجاهات متعددة بدلاً من اتجاه واحد فقط، بحيث قد يكون المُخرج الحالي Current output بمثابة مدخلٍ لما يحدث بعد ذلك. فشبكة البوابات المنطقية Network of logic gates في المعالج الدقيق Microprocessor لحاسوب محمول هي أيضاً شبكات متكررة بهذا المعنى: فهي تواصل إعادة استخدام معلوماتها السابقة، وتسمح بإدخال معلومات جديدة من لوحة المفاتيح ولوحة التتبع والكاميرا وغيرها؛ مما يؤثر في حوسبتها المستمرة، وهذا بدوره يحدد إخراج المعلومات مثلاً على شاشة أو سماعة أو طابعة أو شبكة لاسلكية. وبشكل مشابه، فإن شبكة الخلايا العصبية التي يشغلها الدماغ شبكة متكررة، تسمح بمدخلات من المعلومات من عينيك، في حين تؤثر الحواس الأخرى في حوسبتها المستمرة، التي بدورها تحدد مخرجات المعلومات الموجَّهة إلى عضلاتك.

وتاريخ التعلم هو على الأقل بطول تاريخ الحياة نفسه، لمّا كان كل كائن حي يتكاثر ذاتيّاً يؤدي عملية نسخ ومعالجة معلومات مثيرة للاهتمام - وهو سلوك تَمَّ تعلمه بطريقة ما، لكن، خلال عصر الحياة 1.0، لم تتعلم الكائنات الحية خلال حياتها: فقوانينها لمعالجة المعلومات ولإصدار ردود الفعل كانت محددة من قبل حمضها النووي الذي ورثته، ومن ثم فإن التعلّم الوحيد حدث ببطء على مستوى الأنواع Species، من خلال التطور الداروني عبر الأجيال.

قبل نصف بليون سنة ماضية تقريباً، هنا على الأرض اكتشفت سلالات جينية معينة طريقة لجعل الحيوانات تحتوي على شبكات عصبية، قادرة على تعلم السلوك من التجارب خلال فترة حياتها. وقد وصل الإصدار الثاني من الحياة Life 2.0، وبسبب قدرة هذه الحيوانات على التعلم بشكل أسرع وتفوقها على المنافسة، انتشرت كالنار في الهشيم في جميع أنحاء الكرة الأرضية. وكما استكشفنا في الفصل الأول، صارت الحياة أفضل تدريجيّاً في التعلم، واستمر ذلك بمعدل متزايد. وفي نوع معين يشبه القردة Ape نما دماغ متكيف لاكتساب المعرفة، بحيث تعلم كيفية استخدام الأدوات وإشعال النار والتحدث بلغة وخلق مجتمع عالمي معقد. ويمكن النظر إلى هذا المجتمع نفسه على أنه نظام يتذكر، ويحوسب ويتعلم، كل ذلك بوتيرة متسارعة يُمكّن أثناءها اختراع واحد للاختراع الذي يليه: الكتابة، فالمطبعة، فالعلوم الحديثة، فالحواسيب، فالإنترنت، إلخ. فما الذي سيضعه المؤرخون في المستقبل على قائمة الاختراعات التمكينية؟ تخميني هو الذكاء الاصطناعي.

وكما نعلم جميعاً، فإن التحسينات الهائلة في كل من ذاكرة الحاسوب والقوة الحوسبية (الشكل 4.2 والشكل 8.2) تُرجمت إلى تقدم مذهل في الذكاء الاصطناعي - ولكن نضوج تعلّم الآلة استغرق وقتاً طويلاً. وعندما تغلب الحاسوبُ ديب مايند Deep Mind من الشركة IBM على بطل الشطرنج غاري كاسباروف Garry Kasparov في عام 1997، كانت أهم مميزاته هي الذاكرة والحوسبة، وليس التعلم. فقد طوّر فريق من البشر ذكاءه الحوسبي Computational intelligence، وكان السبب الرئيسي وراء قدرة ديب بلو على التغلب على مطوريه هو قدرته على الحوسبة بشكل أسرع، ومن ثم تحليل المزيد من المواقف المحتملة. وعندما أزاح الحاسوب واتسون Watson من الشركة IBM بطل العالم البشري في مسابقة «جيوباردي!» Jeopardy! عن عرشه، فقد اعتمد أيضاً على التعلم بدرجة أقل من اعتماده على مهارات مبرمجة خصيصاً وذاكرة وسرعة فائقتين. ويمكن إطلاق القول نفسه عن معظم الاختراقات المبكرة في مجال الروبوتات، بدءاً من الحركة على سيقان ذاتية إلى السيارات ذاتية القيادة والصواريخ التي تهبط ذاتيا.

وفي المقابل، كانت القوة الدافعة وراء العديد من أحدث الإنجازات الخارقة للذكاء الاصطناعي هي تعلم الآلة، مثلاً، خذ بعين الاعتبار الشكل 11.2، من السهل عليك أن تخبر ما في الصورة، ولكن لبرمجة دالة لا تُدخل شيئاً سوى ألوان جميع بكسلات الصورة وتنتج تسجيلاً دقيقاً مثل «مجموعة من الشباب يلعبون لعبة الفريزبي» استعصى على جميع الباحثين في الذكاء الاصطناعي بالعالَم لعقود من الزمن، ولكن في عام 2014 تمكن فريق من غوغل Google بقيادة إليا سوتسكيفر Ilya Sutskever من القيام بذلك على وجه التحديد. أَدخِلْ مجموعة مختلفة من ألوان البكسل، وسيجيب الحاسوب «قطيع من الأفيال يسير في حقل عشبي جاف» مرة أخرى إجابة صحيحة، كيف فعلوا ذلك؟ على غرار الحاسوب ديب بلو، عن طريق برمجة الخوارزميات يدويّاً للكشف عن قرص الفرزبي Frisbees والوجوه وما شابه ذلك؟ لا، من خلال إنشاء شبكة عصبية بسيطة نسبيّاً من دون أي معرفة على الإطلاق بالعالم المادي أو محتوياته، ثم تركها تتعلم من خلال

تعريضها لكميات هائلة من البيانات. وفي عام 2004 كتب مُنظِّر الذكاء الاصطناعي جيف هوكينز Jeff Hawkins أن «لا حاسوب... يستطيع أن يرى بنفس مقدرة فأر». ولكن تلك الأيام ولت منذ زمن طويل.



الشكل 11.2: «مجموعة من الشباب يلعبون لعبة الفريزبي» - هذا التعليق كتب بواسطة حاسوب من دون أي فهم للأشخاص أو الألعاب.

وتماماً مثلما لا نفهم تماماً الكيفية التي يتعلم بها أطفالنا، فإننا لا نزال لا نفهم تماماً الكيفية التي تتعلم بها هذه الشبكات العصبية، ولماذا تفشل أحياناً. ولكن ما هو واضح هو أنها مفيدة جدّاً وتستثير موجة في الاستثمارات بالتعلم العميق. فقد حول التعلم العميق الآن العديد من جوانب رؤية الحاسوب Computer vision، من نسخ الكتابة اليدوية إلى تحليل الفيديو في الوقت الحقيقي للسيارات ذاتية القيادة. وكما أحدثت ثورة في قدرة الحواسيب على تحويل لغة منطوقة إلى نص مكتوب وترجمتها إلى لغات أخرى، حتى في الوقت الحقيقي - وهذا هو السبب في أننا نستطيع الآن التحدث إلى المساعدين الرقميين الشخصيين مثل سيري Siri وغوغل ناو Google Now وكورتانا Cortana. وتزداد صعوبة تلك الألغاز المزعجة من اختبارات CAPTCHA إذ نحتاج إلى إقناع موقع ويب بأننا بشر، ولاستباق ما تستطيع تكنولوجيا التعلم الآلة فعله، في عام 2015، أصدر غوغل ديب مايند Google DeepMind نظامَ ذكاءٍ اصطناعي باستخدام التعلم العميق قادراً على إتقان العشرات من ألعاب الحاسوب مثلما يتمكن طفل من ذلك -من دون أي تعليمات على الإطلاق- إلا أنه سرعان ما تعلم اللعب بشكل أفضل من أي إنسان. وفي عام 2016 بنت الشركة نفسها نظام آلفا غو AlphaGo، وهو نظام حاسوب يلعب لعبة غو Go-playing استخدم التعلم العميق لتقييم قوة المحطات المختلفة على اللعبة اللوحية

وهزم أقوى بطل من أبطال لعبة Go في العالَم، وهذا التقدم يغذي حلقة حميدة، إذ يجلب المزيد من التمويل والمواهب إلى أبحاث الذكاء الاصطناعي؛ مما يؤدي إلى مزيد من التقدم.

لقد قضينا هذا الفصل في استكشاف طبيعة الذكاء وتطوره حتى الآن، كم من الوقت سيستغرق الأمر حتى تتمكن الآلات من التغلب علينا في **جميع** المهام المعرفية Cognitive tasks؟ من الواضح أننا لا نعرف، ويجب أن نكون منفتحين على إمكانية أن تكون الإجابة «أبدا». لكن الرسالة الأساسية لهذا الفصل هي أننا نحتاج أيضاً إلى الأخذ بعين الاعتبار إمكانية حدوث ذلك، وربما حتى في حياتنا. ففي المحصلة، يمكن ترتيب المادة بحيث عندما تتبع قوانين الفيزياء، فإنها تتذكر وتحوسب وتتعلم. والمادة لا تحتاج إلى أن تكون بيولوجية، وكثيراً ما كان باحثو الذكاء الاصطناعي متهمين بالإفراط في الوعود وتطوير ما هو أقل من الموعود. ولكن إنصافاً، فإن بعض منتقديهم ليس لديهم سجل أفضل، إذ يستمر البعض بتحريك مكان قوائم شبك الأهداف، فيظل فعليّاً يُعرّف الذكاء على أنه ذلك الشيء الذي لا يزال الحاسوب غير قادر على القيام به، أو يُعرّفه بذلك الذي يثير إعجابنا، والآلات الآن جيدة أو ممتازة في الحساب، والشطرنج، وبرهنة نظرية رياضيات، والانتقاء من مجموعة، وكتابة شرح الصور، والقيادة، ولعب بآلات اللعب، ولعبة غو، وتحليل تركيب الكلام، ونسخ الكلام، والترجمة، وتشخيص السرطان. ولكن بعض النقاد سيسخرون بازدراء «بالتأكيد - لكن هذا ليس بالذكاء الحقيقي!» وقد يستمرون في الجدل بأن الذكاء الحقيقي لا يشمل سوى قمم الجبال في مشهد مورافيك (الشكل 2.2) التي لم تُغمر بعد، تماماً كما اعتاد البعض في الماضي أن يجادلوا في أن إعداد شروح الصور ولعب اللعبة غو هي من الأمور التي تعد ذكاء حقيقيّاً - إلا أن المياه واصلت الارتفاع.

على افتراض أن المياه ستستمر في الارتفاع لفترة أطول على الأقل، فإن تأثير الذكاء الاصطناعي في المجتمع سيستمر في النمو. وقبل وقت طويل من وصول الذكاء الاصطناعي إلى المستوى البشري في جميع المهام، سيعطينا فرصاً ويضعنا أمام تحديات مذهلة فيما يختص بقضايا مثل أخطاء البرمجة Bugs والقوانين والأسلحة والوظائف، ما هي وكيف يمكننا أن نعد لها على أفضل وجه؟ دعونا نستكشف هذا في الفصل التالي.

الخلاصة:

- لا يمكن قياس الذكاء، المُعرَّف بأنه القدرة على تحقيق الأهداف Goals المعقدة، اعتماداً على معدل ذكاء IQ واحد، وإنّما من خلال طيف القدرة Ability spectrum لجميع الأهداف.
- اليوم يميل الذكاء الاصطناعي إلى أن يكون **ضيقاً**، مع كون كل نظام قادراً على تحقيق أهداف محددة جدّاً فقط، في حين أن الذكاء البشري **واسع** بشكل ملحوظ.
- تتميز الذاكرة والحوسبة والتعلم والذكاء بإحساس مجرد وغير ملموس وأثيري؛ لأنها مستقلة عن المادة Substrateindependent: قادرة على أن تنحى منحى خاصّاً بها لا يعتمد على -أو يعكس- تفاصيل المادة الأساسية التي تقوم عليها.

- أي جزء من المادة قد يكون المادة للذاكرة مادام أنه يحتوي على العديد من الحالات المستقرة المختلفة.
- قد تكون أي مادة **كومبيوترونيوم *Computronium***، مادة الحوسبة Computation، طالما أنها تحتوي على وحدات بناء عامة Universal building blocks معينة يمكن دمجها لتنفيذ أي دالة. وتعتبر البوابات NAND والعصبونات مثالين مهمين على «الذرات الحاسوبية» Computational atoms العامة.
- الشبكة العصبية هي مادة قوية **للتعلم**؛ لأنه -ببساطة بإطاعة قوانين الفيزياء- يمكنها إعادة ترتيب نفسها لتصبح أفضل وأفضل في تنفيذ الحوسبات المطلوبة.
- بسبب بساطة قوانين الفيزياء المذهلة؛ لا نهتم نحن البشر إلا بجزء ضئيل من جميع المشكلات الحاسوبية التي يمكن تخيلها، والشبكات العصبية تميل إلى أن تكون جيدة بشكل ملحوظ في حل هذا الجزء الصغير بالتحديد.
- بمجرد أن تتضاعف قوة التكنولوجيا إلى ضعفين، يمكن استخدامها في كثير من الأحيان لتصميم وبناء تكنولوجيا ذات قوة مضاعفة بدورها؛ مما يؤدي إلى مضاعفة القدرة التي تتضاعف كما يُملي قانون مور، إذ انخفضت تكلفة تكنولوجيا المعلومات إلى النصف تقريباً كل عامين لمدة قرن تقريباً؛ مما مكّن من بزوغ عصر المعلومات.
- إذا استمر تقدم الذكاء الاصطناعي - قبل وقت طويل من وصوله إلى مستوى الذكاء البشري في جميع المهارات- فإنه سيمنحنا فرصاً وتحديات مذهلة فيما يختص بقضايا مثل أخطاء البرمجة Bugs والقوانين والأسلحة والوظائف، التي سنستكشفها في الفصل التالي.

الفصل 3

المستقبل القريب: إنجازات خارقة وأخطاء برمجة وقوانين وأسلحة ووظائف

إذا لم نغير الاتجاه قريباً؛ فسينتهي بنا الأمر إلى حيث نحن ذاهبون الآن.

إروين كوري

ماذا يعني أن تكون إنساناً في الوقت الحاضر وفي العصر الحالي؟ مثلاً، ما الذي نُثمِّنه حقّاً في أنفسنا والذي يجعلنا مختلفين عن سائر أشكال الحياة والآلات الأخرى؟ ما الذي يُثمِّنه الآخرون فينا ويجعلهم على استعداد لعرض وظائف علينا؟ مهما كانت إجاباتنا عن هذه الأسئلة في وقت من الأوقات، فمن الواضح أن صعود التكنولوجيا يجب أن يُغيرها تدريجيّاً.

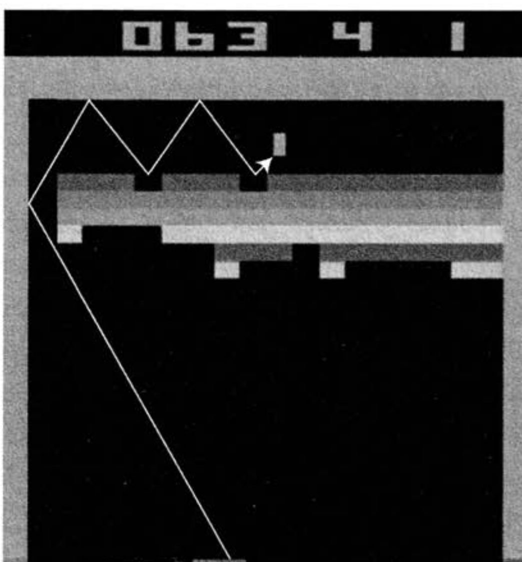
خذني، مثلاً، بصفتي عالِماً، أفتخر بتحديد أهدافي، وباستخدام الإبداع والحدس لمعالجة مجموعة واسعة من المشكلات التي لم تُحلَّ بعد، وباستخدام اللغة لتقاسم ما أكتشفه. لحسن الحظ بالنسبة إليّ، المجتمع مستعد للدفع لي للقيام بذلك كعمل. وقبل قرون، ربما كنت سأقوم، مثل كثيرين آخرين، ببناء هويتي حول كوني مزارعاً أو حرفيّاً. ولكن نمو التكنولوجيا قد قَلَّل من نسبة هذه المهن إلى جزء صغير جدّاً من القوى العاملة. وهذا يعني أنه لم يعد من الممكن للجميع بناء هويتهم حول الزراعة أو الحرف اليدوية.

شخصيّاً، لا يزعجني أن آلات اليوم تفوقني في المهارات اليدوية مثل الحفر والحياكة، لأن هذه ليست هواياتي ولا مصدر رزقي ولا أستقي قيمتي الشخصية منها. وفي الواقع، في سن الثامنة سُحق أي وهم ربما كان لدي حول قدراتي في هذا الصدد، عندما أجبرتني مَدْرَسَتي على حضور مقرر دراسي للحرف الفنية لم يكن لدي أي اهتمام به، ولم أكمل مشروعي إلا بفضل مساعدة رحيمة من طالبة في الصف الخامس أشفقت علي.

ولكن مع استمرار تحسن التكنولوجيا، هل سيؤدي صعود الذكاء الاصطناعي في النهاية إلى التغلب على تلك القدرات التي توفر شعوري الحالي بتقدير الذات والقيمة في سوق العمل؟ أخبرني ستيوارت رسل Stuart Russell أنه وكثير من زملائه الباحثين في الذكاء الاصطناعي وجدوا أنفسهم في لحظة «يا للهول!»، عندما شاهدوا ذكاء اصطناعياً يقوم بشيء لم يتوقعوا أن يروه يحدث إلا بعد عدة سنوات، وبهذه الروح، اسمحوا لي بأن أخبركم عن بعض لحظات «يا للهول!» التي مررت بها، وكيف أراها تُؤْذن بتجاوز إمكانات القدرات البشرية قريباً.

إنجازات خارقة

وكلاء التعلم بالتعزيز العميق Deep Reinforcement Learning Agents واجهت لحظة كبرى من كبريات لحظات فَغْرِ الفاهِ تعجباً في عام 2014 أثناء مشاهدة فيديو النظام الذكاء الاصطناعي ديب مايند DeepMind وهو يتعلم لعبة حاسوبية. على وجه التحديد، كان الذكاء الاصطناعي يلعب لعبة بريك آوت Breakout (انظر: الشكل 1.3)، وهي لعبة أتاري Atari كلاسيكية أتذكرها بمحبة من صباي. الهدف من ذلك هو تحريك المضرب بحيث ترتد الكرة مراراً وتكراراً عن جدار من الطوب؛ في كل مرة تصيب الضربة لَبِنَة من الطوب، تختفي اللبنة وتزداد درجاتك.



الشكل 1.3: بعد تعلم لعب لعبة أتاري بريك آوت Atari Breakout من الصفر، باستخدام التعلم بالتعزيز العميق Deep reinforcement learning لتحقيق أقصى الدرجات، اكتشف الذكاء الاصطناعي ديب مايند DeepMind الاستراتيجية المثالية: حفرُ ثقبٍ في أقصى اليسار من جدار الطوب وترك الكرة ترتد عن الجدار من الخلف، فتُجمع النقاط بسرعة كبيرة. الأسهم في الشكل توضح المسارات السابقة للكرة وتحريك المضرب.

لقد كتبت بعض ألعاب الحاسوب في أيام صباي، وكنت أدرك جيداً أنه لم يكن من الصعب كتابة برنامج يمكنه لعب لعبة بريك آوت، لكن هذا ليس هو ما فعله فريق ديب مايند. وبدلاً من ذلك، قاموا بتطوير ذكاء اصطناعي عبارة عن صفحة فارغة لا تعرف شيئاً عن هذه اللعبة - أو عن أي لعبة أخرى، أو حتى عن **مفاهيم** مثل الألعاب أو المضارب أو الطوب أو الكرات. وجلُّ ما عرفه ذكاؤهم الاصطناعي هو أنه تَمَّ إدخال قائمة طويلة من الأرقام فيها فترات منتظمة: النتيجة الحالية وقائمة طويلة من الأرقام التي *سنتعرف* عليها (ولكن الذكاء الاصطناعي لا يعرف ذلك) كمواصفات لكيفية تلوين أجزاء مختلفة من الشاشة. وببساطة أُخبر الذكاء الاصطناعي بزيادة النتيجة عن طريق مخرجات تُخرج، على فترات زمنية منتظمة، أرقاماً *سنتعرف* عليها (ولكن الذكاء الاصطناعي لا يعرف ذلك) ككودات Codes لأي من المفاتيح يجب الضغط عليها.

وفي البداية، لعب الذكاء الاصطناعي بشكل مريع: كان يهز المضرب ذهاباً وإياباً على ما يبدو بشكل عشوائي ولا يصيب الكرة في كل مرة تقريباً. وبعد فترة من الوقت، بدا كما لو أن الذكاء الاصطناعي فهم أن فكرة تحريك المضرب نحو الكرة كانت فكرة جيدة، على الرغم من أنه كان يخطئ إصابتها معظم الوقت. ولكنه استمر بالتحسن مع الممارسة، وسرعان ما صار أفضل في اللعبة مني أنا، فيصيب الكرة بنجاح بغض النظر عن مدى سرعة اقترابها. بعد ذلك فغرتُ فاهي مدهوشاً: لقد نجح في تطوير استراتيجية مدهشة لتحقيق أقصى قدر من الدرجات بالتصويب دائماً إلى الركن الأيسر الأعلى من الجدار وترك الكرة تنحشر هناك مرتطمة بين ظهر الجدار والحائط خلفه، وهذا يبدو كأنه حلٌّ ذكي حقّاً. وفي الواقع، أخبرني ديميس هاسابيس Demis Hassabis لاحقاً بأن المبرمجين في فريق ديب مايند لم يعرفوا هذه الحيلة حتى تعلموها من الذكاء الاصطناعي الذي طوّروه. أنصح بمشاهدة فيديو لهذا بنفسك على الرابط الموجود في الهوامش.[1]

وكانت هناك سمة تشبه السمات الإنسانية؛ لهذا وجدتها مقلقة إلى حد ما: كنت أشاهد ذكاءً اصطناعيّاً كان له هدف، وتعلّم أن يستمر في تحسين تحقيق ذلك الهدف، وفي النهاية يتفوق على مبتدعيه. وفي الفصل السابق، عرّفنا الذكاء بأنه ببساطة القدرة على تحقيق الأهداف المعقدة، لذلك في هذا السياق، كان الذكاء الاصطناعي ديب مايند ينمو أكثر ذكاءً أمام عيني (وإن كان ذلك بالمعنى الضيق جدّاً للعب هذه اللعبة بالذات). وفي الفصل الأول صادفنا ما يسميه علماء الحاسوب **العوامل الذكية**: الكيانات التي تجمع المعلومات حول بيئتها من المستشعرات ثم تعالج هذه المعلومات لتحديد كيفية العمل على بيئتها. وعلى الرغم من أن ذكاء ديب مايند الاصطناعي الذي يلعب الألعاب الحاسوبية يعيش في عالم افتراضي بسيط جدّاً يتكون من الطوب والمضارب والكرات، فإنني لا أنكر أنه عميل ذكي.

سرعان ما نشر فريق ديب مايند طريقتهم ونشروا الكود Code الخاص بها، موضحين أنه يستخدم فكرة بسيطة جدّاً ولكنها قوية تُسمى التعلم بالتعزيز العميق.[2] **والتعلم بالتعزيز العميق** ***Deep reinforcement learning*** هو تقنية التعلم الآلي الكلاسيكي المستوحاة من

علم النفس السلوكي، إذ إن الحصول على مكافأة موجبة يزيد من ميلك إلى فعل شيء من جديد والعكس صحيح. تماماً مثلما يتعلم الكلب ممارسة الحيل عندما يزيد هذا من احتمال حصوله على التشجيع أو تناول وجبة خفيفة من مالكه قريباً. تعلم ذكاء ديب مايند الاصطناعي تحريك المضرب للقبض على الكرة؛ لأن هذا زاد من احتمال حصوله على المزيد من النقاط سريعاً. بعد ذلك دمج فريق ديب مايند هذه الفكرة مع التعلم العميق: فقد درّبوا شبكة عصبية عميقة، كما في الفصل السابق، للتنبؤ بعدد النقاط التي سيتم اكتسابها في المتوسط عند الضغط على كل مفتاح مسموح به على لوحة المفاتيح. بعد ذلك حدد الذكاء الاصطناعي أي مفتاح حدّدته الشبكة العصبية على أنه الأكثر وعداً بالنظر إلى الوضع الحالي للّعبة.

عندما أدرجت السمات التي تساهم في شعوري الشخصي بتقدير الذات كإنسان، فقد شملت القدرة على معالجة مجموعة واسعة من المشكلات التي لم تُحل بعد. وفي المقابل، فإن القدرة على لعب لعبة بريك آوت وعدم القيام بأي شيء آخر سوى ذلك يشكل ذكاءً ضيقاً جدّاً. وفي رأيي، الأهمية الحقيقية للإنجاز الخارق لديب مايند هي أن التعلم بالتعزيز العميق هو أسلوب عام تماماً. ومن المؤكد أنه سُمح للذكاء الاصطناعي نفسه بممارسة 49 لعبة مختلفة من ألعاب أتاري، وتعلّم التغلب على مختبريهم من البشر أي على 29 منهم، من لعبة بونغ Pong إلى بوكسينغ Boxing، وفيديو بينبول Video Pinball وغزاة الفضاء Space Invaders.

لم يستغرق الأمر وقتاً طويلاً حتى بدأت فكرة الذكاء الاصطناعي نفسها بإثبات نفسها في ألعاب أكثر حداثة كانت عوالمها ثلاثية الأبعاد وليست ثنائية الأبعاد. أصدرت شركة أوبن آي OpenAI في سان فرانسيسكو - وهم منافسو شركة ديب مايند- منصة تسمى يونيفيرس (الكون) Universe، حيث يمكن لذكاء ديب مايند الاصطناعي وغيره من العملاء الأذكياء التفاعل مع جهاز حاسوب بكامله كما لو كانوا لعبة: النقر على أي شيء، وكتابة أي شيء، وفتح تشغيل أي برنامج يمكنهم التنقل عبره - باكتشافهم لمتصفح الويب Web browser بالارتباط بمتصفح الويب والبحث عبر الإنترنت، مثلاً.

وبالنظر إلى مستقبل التعلم بالتعزيز العميق وتحسينه، لا توجد نهاية واضحة في الأفق. فلا تقتصر الإمكانات على عوالِم الألعاب الافتراضية، لأنه إذا كنت روبوتاً؛ فيمكن اعتبار الحياة نفسها لعبة. أخبرني ستيوارت راسل أن أول لحظة عظيمة من لحظات «يا للهول!» كانت عندما شاهد الروبوت بيغ دوغ (الكلب الكبير) Big Dog يركض على منحدر مغطى بالثلوج، ويحل بأناقة مشكلة الخطوة الواثقة على أقدام، وهي المسألة التي كان هو نفسه قد ناضل من أجل حلها لسنوات عديدة.[3] ولكن الوصول إلى هذه المرحلة الرئيسية في عام 2008 كان بفضل قدر كبير من العمل من قبل المبرمجين الأذكياء، بعد إنجاز ديب مايند الخارق. لا يوجد أي سبب يمنع الروبوت من استخدامِ نوع Variant من التعزيز العميق في نهاية المطاف لتعليم نفسه المشي دون مساعدة من المبرمجين البشريين: كل ما نحتاج إليه هو نظام يعطي نقاطاً كلما أحرز البرنامج تقدماً. وبالمثل، لدى الروبوتات في العالم

الواقعي القدرة على تعلم السباحة، ولعب كرة تنس الطاولة، وأداء ما لا نهاية له من المهام الحركية الأخرى من دون مساعدة من المبرمجين البشر. ولتسريع الأمور ولتقليل خطر الانحشار في مكان ما أو الإصابة بضرر أثناء عملية التعلم، فربما يلجؤون إلى الواقع الافتراضي في المراحل الأولى من تعلمهم.مكتبة سُر مَن قرأ

الحدس والإبداع والاستراتيجية

في اعتقادي أن اللحظة التي فاز فيها نظام الذكاء الاصطناعي ألفا غو AlphaGo من ديب مايند بمباراة Go الأولى ضد لي سيدول Lee Sedol، الذي يعتبر أفضل لاعب في العالم في أوائل القرن العشرين، هي لحظة مهمة أخرى.

كان من المتوقع أن تزيح الآلات لاعبي اللعبة غو عن عروشهم في مرحلة ما، كما حصل مع زملائهم من لاعبي الشطرنج قبل عقدين من الزمن. ولكن، معظم خبراء اللعبة غو توقعوا أن الأمر سيستغرق عقداً آخر، لذلك كان انتصار النظام ألفا غو لحظة محورية بالنسبة إليهم، وكذلك بالنسبة إلي. فقد أكدَّ كل من نيك بوستروم Nick Bostrom وراي كورزويل على مدى صعوبة التنبؤ بموعد تحقيق إنجازات خارقة في الذكاء الاصطناعي، وهو ما يتضح من المقابلات التي أجراها مع لي سيدول نفسه قبل وبعد الخسارة في أول ثلاث مباريات:

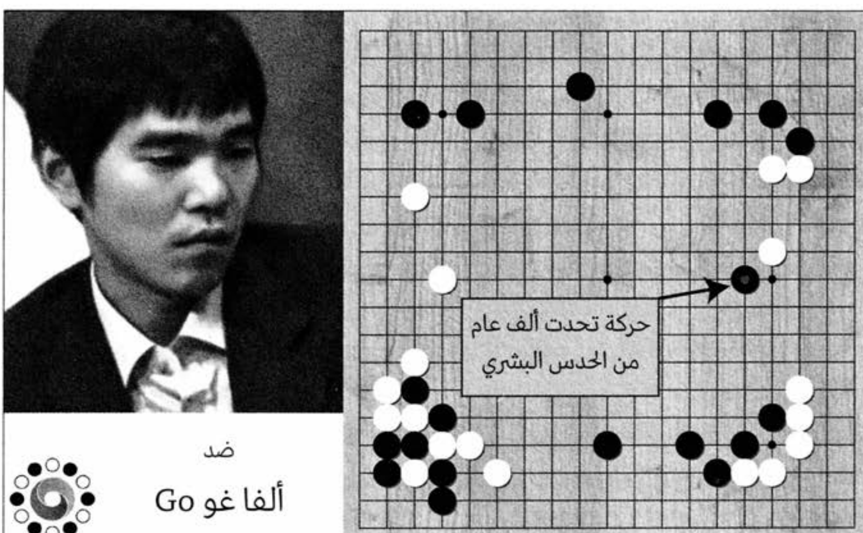
- أكتوبر 2015: «بناءً على مستواه المشاهد... أعتقد أنني سأفوز بالمباراة فوزاً ساحقاً بسرعة».
- فبراير 2016: «سمعت أن ديب مايند الذكاء الاصطناعي من غوغل Google قوي بشكل مذهل ويزداد قوة، لكنني متأكد من أنني أستطيع الفوز على الأقل هذه المرة».
- 9 مارس 2016: «لقد فوجئت جدّاً؛ لأنني لم أعتقد أنني سأخسر».
- 10 مارس 2016: «أنا عاجز عن الكلام تماماً،، أنا في صدمة، أنا أعترف بذلك،، المباراة الثالثة لن تكون سهلة عليّ».
- 12 مارس 2016: «شعرت بالعجز».

في غضون عام بعدما لعب ضد لي سيدول، لعب نظام ألفا غو AlphaGo المُحسّن ضد أفضل 20 لاعباً في العالم جميعهم دون أن يخسر مباراة واحدة.

لماذا كان هذا أمراً مهمّاً في تقديري؟ حسناً، اعترفت آنفاً بأنني أرى الحدس والإبداع عنصرين من سماتي الإنسانية الأساسية. وكما أوضح في هذا المقام، أشعر بأن النظام آلفا غو AlphaGo عرض كلتا السمتين.

يتناوب لاعبو اللعبة غو Go في وضع الحجارة البيضاء والسوداء على الرقعة 19 × 19 (انظر الشكل: 2.3). وعدد مواضع Go الممكنة عدد ضخم لدرجة أنه أكبر بكثير من

عدد الذرات في عالمنا؛ مما يعني أن محاولة تحليل كل المتواليات المثيرة للاهتمام من الحركات المستقبلية ستفشل بسرعة. لذلك يعتمد اللاعبون اعتماداً كبيراً على الحدس اللاواعي لدعم تفكيرهم الواعي. ويُطوِّر الخبراء شعوراً غامضاً بأماكن المواضع القوية وتلك التي تكون ضعيفة. وكما رأينا في الفصل الأخير، تكون نتائج التعلم العميق في بعض الأحيان تشبه الحدس شبهاً بعيداً: قد تحدد الشبكة العصبية العميقة أن الصورة تصور قطة دون أن تكون قادرة على شرح السبب. لذلك راهن فريق ديب مايند على فكرة أن التعلم العميق قد يكون قادراً على التعرف ليس فقط على القطط، ولكن أيضاً على مواضع Go القوية. وكانت الفكرة الأساسية التي قاموا ببنائها في النظام AlphaGo هي المزاوجة من القوة البديهية للتعلم العميق مع القوة المنطقية للنظم GOFAI - التي تمثل اختصاراً لما يعرف بـ "الذكاء الاصطناعي التقليدي العادي" Good Old-Fashioned AI مما قبل ثورة التعلم العميق. واستخدموا قاعدة بيانات ضخمة لمواقع Go من لعب الأفراد واللعبات التي لعب فيها النظام AlphaGo ضد نسخة من نفسه. ودربوا شبكة عصبية عميقة للتنبؤ من كل موضع باحتمال فوز الأبيض White في النهاية. كما درّبوا شبكة منفصلة للتنبؤ بالتحركات المحتملة القادمة. بعد ذلك دمجوا هاتين الشبكتين الجديدتين مع المنهج GOFAI الذي بحث بذكاء خلال قائمة مجردة من تسلسل الحركات المستقبلية المحتملة لتحديد الخطوة التالية التي من شأنها أن تؤدي إلى أقوى موضع في نهاية اللعبة.



الشكل 2.3: قام AlphaGo AI من ديب مايند بخطوة مبدعة جدّاً على الخط 5، هازماً آلاف السنين من الحكمة الإنسانية، الخطوة التي أثبتت بعد نحو 50 حركة أهميتها لهزيمة لي سيدول Lee Lee Sedol أسطورة اللعبة غو Go.

هذه المزاوجة بين الحدس والمنطق ولدت حركات لم تكن مجرد حركات قوية، ولكن في بعض الأحيان إبداعية جدّاً. مثلاً، تملي آلاف السنين من حكمة اللعبة Go أنه من الأفضل في بداية اللعبة على الخط الثالث أو الرابع من الحافة، هناك مقايضة بين الطريقتين: اللعب على الخط الثالث يساعد على تحقيق مكاسب على المدى القصير في اتجاه الطرف الآخر من اللوحة، في حين أن اللعب على الرابع يساعد تنفيذ التأثير الاستراتيجي طويل المدى على التحرك نحو الوسط.

في الخطوة السابعة والثلاثين من اللعبة الثانية، صدم النظام AlphaGo عالَم اللعبة غو Go بتحدي تلك الحكمة القديمة واللعب على السطر الخامس (الشكل 2.3)، كما لو كان النظام أكثر ثقة من البشر من حيث قدراته التخطيطية طويلة الأجل. ولذلك فضّل الميزة الاستراتيجية على المكاسب قصيرة الأجل، وقد صُعق المعلقون، بل إن لي سيدول نهض وغادر الغرفة لبرهة.[4] وبالتأكيد بعد نحو 50 خطوة في وقت لاحق، انتهى الأمر بالركض من الركن الأيسر السفلي من اللوحة بالانتشار عبر رقعة اللعب والوصول إلى ذلك الحجر الأسود من الحركة السابعة والثلاثين! وهذا الحافز هو ما ربح اللعبة في نهاية المطاف؛ مما عزّز إرث الخطوة من الصف الخامس في النظام AlphaGo باعتبارها واحدة من أكثر الألعاب إبداعاً في تاريخ اللعبة غو Go.

ونظراً لجوانبها البديهية والإبداعية، يُنظر إلى اللعبة Go على أنها شكل فني أكثر منها مجرد لعبة أخرى، فقد كانت تُعتبر واحدةً من «الفنون الأساسية» الأربعة في الصين القديمة، إلى جانب الرسم والخط وموسيقى تشين، وهي لا تزال تحظى بشعبية كبيرة في آسيا. وقد شاهد اللعبة الأولى بين النظام AlphaGo ولي سيدول ما يقرب من 300 مليون شخص. ونتيجة لذلك اهتز العالمَ هزة قوية، واعتبر فوز ألفا غو AlphaGo مرحلة مهمة للبشرية. وهذا ما قاله Ke Jie، لاعب غو Go الأول على مستوى العالم في ذلك الوقت[5]: «لقد لعبت الإنسانية لعبة غو Go لآلاف السنين، لكن، كما بَيّن لنا الذكاء الاصطناعي، لم نخدش السطح بعد... اتحاد اللاعبين من البشر والحاسوب سيأذنون بدخول حقبة جديدة... معا سيتمكّن الإنسان والذكاء الاصطناعي من معرفة حقيقة اللعبة Go. ومثال هذا التعاون المثمر بين الإنسان والآلة يبدو واعداً بالفعل في العديد من المجالات، بما في ذلك العلوم، إذ من المأمول أن يساعدنا -نحن البشر- الذكاءُ الاصطناعي على تعميق فهمنا وتحقيق إمكاناتنا النهائية.

وأعتقد أن النظام ألفا غو AlphaGo سيعلمنا درساً آخر مهمّاً في المستقبل القريب: قد تنتج منه استراتيجية لا مثيل لها من الجمع بين الحدس في التعلم العميق مع منطق نظم الذكاء الاصطناعي GOFAI Go. ولما كانت اللعبة غو Go واحدة من أقصى ألعاب **الاستراتيجية**، فإن الذكاء الاصطناعي مهيأ الآن للتخرج (من التدريب) ولتحدي (أو مساعدة) أفضل الاستراتيجيين من البشر حتى إلى ما هو أبعد من رُقع الألعاب - مثلاً الاستراتيجية الاستثمارية، والاستراتيجية السياسية، والاستراتيجية العسكرية. وعادةً ما تكون مشكلات الاستراتيجية في العالم الواقعي هذه مُعقَّدة بفعل السيكولوجية البشرية، والمعلومات الناقصة والعوامل التي يجب أن تُنمذج على أنها عوامل عشوائية. غير أن أنظمة الذكاء الاصطناعي التي تلعب البوكر قد أظهرت بالفعل أنه لا يمكن التغلب على أي من هذه التحديات.

اللغة الطبيعية

هناك مجال آخر أذهلني فيه تقدم الذكاء الاصطناعي مؤخراً. فقد أحببت السفر مبكراً في الحياة، وشكَّل فضولي للثقافات واللغات الأخرى جزءاً مهمّاً من هويتي، فقد نشأت وأنا أتحدث اللغة السويدية والإنجليزية، وقد درستُ اللغة الألمانية والإسبانية في المدرسة، وتعلمتُ البرتغالية والرومانية من خلال زواجين، وعلّمت نفسي بعضاً من الروسية والفرنسية والماندرين من أجل المتعة.

غير أن الذكاء الاصطناعي آخذ بالامتداد، وبعد اكتشاف مهم في عام *2016*، لا توجد تقريباً لغات برمجة كسولة *Lazy languages* يمكنني أن أترجم ما بينها أفضل من نظام الذكاء الاصطناعي الذي طوره فريق دماغ غوغل*Google*.

هل أَفْصَحتُ عما أريد قوله بوضوح؟ هذا ما كنت أحاول بالفعل أن أقوله: **غير أن الذكاء الاصطناعي آخذ بالتطور. وبعد إنجاز خارق في عام *2016*، لم يتبقَّ من اللغات تقريباً ما يمكنني الترجمة بينها أفضل من نظام الذكاء الاصطناعي الذي طوره فريق غوغل برين (دماغ غوغل) *Google Brain*.**

ولكنني ترجمت النص أولاً إلى الإسبانية باستخدام تطبيق App حمّلته على حاسوبي المحمول قبل بضع سنوات. وفي عام 2016 رقّى Upgraded غوغل برين Google Brain خدمات الترجمة المجانية من غوغل Google Translate، بحيث يستخدم الشبكات العصبية العميقة المتكررة، والتحسين مقارنة بالنظام GOFAI الأقدم كما ترون هو ضخم جدّاً:[6]

غير أن الذكاء الاصطناعي صار قادراً على مجاراتي، وبعد الإنجاز الخارق في *2016*، هناك بالكاد أي لغات يمكنني الترجمة ما بينها أفضل من نظام الذكاء الاصطناعي الذي طوره فريق غوغل برين *Google Brain*.

كما ترون، فُقد ضمير «ي» خلال المسار الإسباني؛ مما غيَّر مع الأسف المعنى، مقارب، ولكن لم ينجح في نقل المعنى! ولكن، دفاعاً عن الذكاء غوغل الاصطناعي، كثيراً ما أُنتقد لكتابة جمل طويلة من دون داعٍ يصعب تحليلها، وقد اخترت إحدى جملتيَّ الأكثر تشويشاً في هذا المثال. وفي حالة الجمل الأكثر تقليدية يعالج ذكاؤهم الاصطناعي الجمل دون أي خطأ. ونتيجة لذلك أثارت ضجة كبيرة عندما أُطلق أول مرة. وهو مفيد بما فيه الكفاية لاستخدام مئات الملايين من الأشخاص يومياً. إضافة إلى ذلك، بفضل التقدم المحرز مؤخراً في التعلم العميق لتحويل الكلام إلى نص Speech-to-Text وتحويل النص إلى كلام Text-to-Speech، يمكن لهؤلاء المستخدمين التحدث إلى هواتفهم الذكية بلغة ما والاستماع إلى النتيجة المترجمة بلغة أخرى.

ومعالجة اللغة الطبيعية Natural language هي الآن واحد من أسرع المجالات تطوراً في الذكاء الاصطناعي. وأعتقد أن المزيد من النجاح سيكون له تأثير كبير؛ لأن اللغة هي أمر أساسي للإنسان. وكلما تحسّنت قدرات التنبؤ اللغوية لدى الذكاء الاصطناعي، صار

أفضل في إنشاء ردود بريد إلكتروني منطقية أو الاسترسال في محادثة منطوقة. وبالنسبة إلى المراقب الخارجي، فقد يعطي هذا انطباعاً من حدوث عمليات تفكير بشرية. وهكذا، تخطو أنظمة التعلّم العميق خطواتٍ صغيرةً نحو اجتياز اختبار تورينغ الشهير، إذ يتعين على الآلة أن تتحدث جيداً بما يكفي كتابةً لخداع شخص فيعتقد أنه يحاور إنساناً آخر.

ولكن لا يزال الطريق طويلاً أمام الوصول إلى ذكاء اصطناعي لمعالجة اللغة Language-processing AI. وعلى الرغم من أنني يجب أن أعترف بأنني أشعر ببعض القلق عندما يتفوق على ذكاء اصطناعي في الترجمة، فإنني أشعر بارتياح عندما أُذكِّر نفسي أنه، حتى الآن، لا يفهم ما يقوله ولا يفقه منه شيئاً. فمن خلال التدريب على مجموعات ضخمة من البيانات، يكتشف الذكاء الاصطناعي الأنماط والعلاقات بين الكلمات من دون ربط هذه الكلمات بأي شيء في العالم الحقيقي. مثلاً، قد يمثل كل كلمة بقائمة من آلاف الأعداد التي تحدد مدى تشابهها مع بعض الكلمات الأخرى. وقد يستنتج من هذا أن الفرق بين «الملك» و «الملكة» يشبه الفرق بين «الزوج» و «الزوجة» - ولكن لا يزال يفتقد أدنى فكرة عما يعنيه الكون ذكراً أو أنثى، أو حتى إذا كان هناك بالفعل مثل هذا في الواقع المادي من الفضاء والزمن والمادة.

ولما كان اختبار تورينغ يتعلق أساساً بالخداع، فقد انتُقِد لاختباره السذاجة البشرية أكثر منه للذكاء الحقيقي. وعلى النقيض من ذلك، فإن اختباراً منافساً يسمى **تحدي مخطط فينوغراد *Winograd Schema Challenge*** يركِّز على الحدس المنطقي Commonsense الذي تفتقر إليه أنظمة التعلم العميق الحالية. نحن البشر نستخدم بشكل روتيني معرفة العالمَ الحقيقي عند تحليل جملة ما، لمعرفة ما يشير إليه الضمير، مثلاً، يسأل تحدي فينوغراد النموذجي ما الذي يشير إليه الضمير «هم» هنا:

1. «رفض أعضاء مجلس المدينة منح المتظاهرين تصريحاً؛ لأنهم يخشون العنف».
2. «رفض أعضاء مجلس المدينة منح المتظاهرين تصريحاً؛ لأنهم يدعون إلى العنف».

هناك مسابقة سنوية للذكاء الاصطناعي للإجابة عن مثل هذه الأسئلة، ولا يزال أداء الذكاء الاصطناعي سيئاً.[7] وهذا التحدي المُحدَّد، فهم ما الذي يشير إلى ماذا، نسف حتى غوغل ترانسليت GoogleTranslate عندما استعضت عنه الإسبانية بالصينية في مثالي المذكور آنفاً:

لكن الذكاء الاصطناعي لحق بي، بعد إنجاز خارق في عام 2016، مع لا لغة تقريباً، يمكنني ترجمة نظام الذكاء الاصطناعي أكثر مما هو مطور من فريق *Google Brain*.

يرجى تجربة ذلك بنفسك على https://translate.google.com الآن وأنت تقرأ الكتاب، لترى ما إذا كان الذكاء الاصطناعي من غوغل قد تحسن! فهناك احتمال جيد أن يكون قد تحسن بالفعل، نظراً لوجود مقاربات واعدة للمزاوجة بين شبكات عصبية عميقة متكررة بالنظام GOFAI لبناء ذكاء اصطناعي لمعالجة اللغة يتضمن نمذجة عالمية.

فرص وتحديات

من الواضح أن هذه الأمثلة الثلاثة كانت مجرد عينة؛ نظراً لأن الذكاء الاصطناعي يتقدم بسرعة عبر العديد من الجبهات المهمة. إضافة إلى ذلك، على الرغم من أنني ذكرت شركتين فقط في هذه الأمثلة، فإن المجموعات البحثية المتنافسة في الجامعات وغيرها من الشركات في الغالب غير متأخرة عن الركب كثيراً، فيمكن سماع ضجيج الامتصاص بصوت عالٍ في أقسام علوم الحاسوب في جميع أنحاء العالم، إذ تقدم آبل Apple وبايدو Baidu و ديب مايند وفيسبوك Facebook وغوغل وميكروسوفت عروضاً مغرية لشفط الطلبة وزملاء ما بعد الدكتوراه Postdocs وأعضاء هيئة التدريس.

ومن المهم ألا تُضلِّلك الأمثلة التي قدَّمتُها عند عرض تاريخ الذكاء الاصطناعي؛ فتظن أنها فترات ركود تتخللها الإنجازات الخارقة عرضيّاً. ومن موقعي المتميز رأيت بدلاً من ذلك تقدماً ثابتاً إلى حد ما لفترة طويلة - وهو ما تنقله وسائل الإعلام على أنه إنجاز خارق عندما تتجاوز عتبة تُمكِّن تطبيقاً جديداً من الاستحواذ على المُخيِّلة أو تقدِّم منتجاً مفيداً. لذلك، أعتقد أن التقدم السريع للذكاء الاصطناعي ربما يستمر لسنوات عديدة. إضافة إلى ذلك، كما رأينا في الفصل الماضي، لا يوجد أي سبب جوهري يمنع استمرار هذا التقدم حتى يُماثل الذكاء الاصطناعي القدراتِ البشرية في معظم المهام.

ومما يثير السؤال: كيف سيؤثر هذا فينا؟ كيف سيتغير تقدم الذكاء الاصطناعي على المدى القريب ما يعنيه الكون إنساناً؟ فقد رأينا أن الأهداف أو الاتساع أو الحدس أو الإبداع أو اللغة - وهي سمات يشعر الكثيرون بأنها أساسية للكون إنسانا. هذا يعني أنه حتى في المدى القريب، وقبل أن يتمكن أي ذكاء AGI (الذكاء الاصطناعي العام، Artificial general intelligence) من مضاهاتنا في جميع المهام بوقت طويل، قد يكون للذكاء الاصطناعي تأثير كبير في نظرتنا إلى أنفسنا، وفيما يمكننا القيام به عندما يُكمِّلنا ذكاء اصطناعي وعلى ما يمكننا كسبه من مال عندما نتنافس ضد ذكاء اصطناعي، هل سيكون هذا التأثير إلى الأفضل أم الأسوأ؟ وما هي الفرص والتحديات نتيجة ذلك على المدى القريب؟

كل ما نحبه في الحضارة هو نتاج ذكاء الإنسان، لذلك إذا استطعنا تضخيمه باستخدام الذكاء الاصطناعي، فمن الواضح أن لدينا القدرة على جعل الحياة أفضل. حتى التقدم المتواضع في الذكاء الاصطناعي قد يترجم إلى تحسينات كبيرة في العلوم والتكنولوجيا وما يقابلها من تخفيضات في الحوادث والمرض والظلم والحرب والجراح والفقر. ولكن من أجل جني هذه الفوائد من الذكاء الاصطناعي دون خلق مشكلات جديدة، نحتاج إلى الإجابة عن العديد من الأسئلة المهمة، فمثلاً:

1. كيف يمكننا أن نجعل أنظمة الذكاء الاصطناعي في المستقبل أكثر قوة من الأنظمة الحالية، بحيث تفعل ما نريد دون عطل أو خطأ في التشغيل أو اختراق؟

2. كيف يمكننا تحديث أنظمتنا القانونية لتكون أكثر عدالة وفاعلية ومواكَبة للمشهد الرقمي المتغير بسرعة؟
3. كيف يمكننا أن نجعل الأسلحة أكثر ذكاءً وأقل عرضة لقتل المدنيين الأبرياء دون إطلاق سباق تسلح خارج نطاق السيطرة في أسلحة مميتة ذاتية السيطرة؟
4. كيف يمكننا تنمية ازدهارنا من خلال الأتمتة دون حرمان الأفراد من الدخل أو الهدف في الحياة؟

دعنا نكرس بقية هذا الفصل لاستكشاف كل سؤال من هذه الأسئلة على حدة. هذه الأسئلة الأربعة قصيرة الأجل موجهة أساساً إلى علماء الحاسوب وعلماء القانون والاستراتيجيين العسكريين والاقتصاديين على التوالي. ولكن، للمساعدة على الحصول على الإجابات التي نحتاج إليها بحلول الوقت الذي نحتاج إليها فيه، يجب على الجميع الانضمام إلى هذه المحادثة، لأن التحديات -كما سنرى- تتجاوز جميع الحدود التقليدية، سواء بين التخصصات أم بين الدول.

أخطاء البرمجة مقابل ذكاء اصطناعي متين

كان لتكنولوجيا المعلومات بالفعل تأثير إيجابي كبير في كل قطاع من مشاريعنا البشرية فعليّاً، بدءاً من العلوم إلى التمويل والتصنيع والنقل والرعاية الصحية والطاقة والاتصالات. وهذا التأثير يتضاءل مقارنةً بالتقدم المحتمل الذي قد يجلبه الذكاء الاصطناعي، ولكن كلما زاد اعتمادنا على التكنولوجيا، صارت التكنولوجيا أكثر متانة وجدارة بالثقة، للقيام بما نريده لها أن تقوم به.

وعلى مدار تاريخ البشرية، اعتمدنا على المقاربة المجرَّبة والحقيقية نفسها للحفاظ على تكنولوجيا مفيدة: التعلم من الأخطاء، اخترعنا النار، ووقعنا في أخطاء مراراً، ثم اخترعنا مطفأة الحريق، ومخرج الحريق، وإنذار النار، والمطافي. واخترعنا السيارة، وتصادمنا مراراً، ثم اخترعنا أحزمة الأمان، والوسادات الهوائية، والسيارات ذاتية القيادة. حتى الآن، تسببت تقنياتنا عادة في وقوع عدد قليل ومحدود من الحوادث فلم يفُقْ ضررُها فوائدَها. لكن، مع تطويرنا لتكنولوجيات أكثر قوة من أي وقت مضى، فإننا سنصل حتماً إلى نقطة يكون عندها وقوع حادث واحد فقط مدمراً بما فيه الكفاية يفوق جميع الفوائد. فيجادل البعض في أن الحرب النووية العالمية العرضية ستكون مثل ذلك المثال، ويجادل البعض الآخر في أن وباء هندسة حيوية قد يكون مؤهلاً لأن يكون مثل هذا المثال. وفي الفصل التالي، سنستكشف الجدل الدائر حول ما إذا كان الذكاء الاصطناعي قد يسبب في المستقبل انقراض البشر. ولكننا لا نحتاج إلى التفكير في مثل هذه الأمثلة المتطرفة للتوصل إلى الاستنتاج الحاسم: مع تزايد قوة التكنولوجيا، يجب أن نعتمد بدرجة أقل على مقاربة التجربة والخطأ في هندسة السلامة. وبعبارة أخرى،

يجب أن نحرص على الفعل أكثر من ردة فعل، فنستثمر في أبحاث السلامة التي تهدف إلى منع وقوع الحوادث ولو مرة واحدة، هذا هو السبب في أن المجتمع يستثمر في سلامة المفاعل النووي أكثر من سلامة مصيدة الفئران.

وهذا هو السبب أيضاً، كما رأينا في الفصل الأول. وفي وجود اهتمام شديد من المجتمع بأبحاث السلامة في مجال الذكاء الاصطناعي في مؤتمر بورتوريكو، فقد تعطلت أجهزة الحاسوب وأنظمة الذكاء الاصطناعي دائماً، ولكن هذه المرة ستكون مختلفة: الذكاء الاصطناعي آخذ بالدخول تدريجيّاً في العالمَ الواقعي، ولن يكون مجرد مصدر إزعاج إذا تعطلت شبكة الكهرباء أو سوق الأسهم أو نظام الأسلحة النووية. وفي بقية هذا القسم، أودُّ أن أقدمكم إلى المجالات الرئيسية الأربعة للبحث الفني في مجال سلامة الذكاء الاصطناعي التي تهيمن على دراسة سلامة الذكاء الاصطناعي حاليّاً والتي يتم متابعتها في جميع أنحاء العالم: التحقق Verification، والتدقيق Validation، والأمن Security، والتحكم Control.* وللحيولة دون أن تكون الأمور تقنية وجافة جدّاً، دعونا نفعل ذلك من خلال استكشاف النجاحات السابقة وإخفاقات تكنولوجيا المعلومات في مجالات مختلفة، إضافة إلى الدروس القيمة التي يمكننا أن نتعلم منها والبحث عن التحديات التي تطرحها.

وعلى الرغم من أن معظم هذه القصص قديمة، وتتضمن أنظمة حواسب منخفضة التكنولوجيا لا يشير إليها أي شخص تقريباً على أنها ذكاء اصطناعي، والتي تسببت في عدد قليل من الحوادث هذا إذا تسبب في أي منها، فإننا سنرى أنها على الرغم من ذلك تعلمنا دروساً قيّمة في كيفية تصميم أنظمة ذكاء اصطناعي قوية -بطرق آمنة في المستقبل- قد تكون إخفاقاتها كارثية حقّاً.

ذكاء اصطناعي لاستكشاف الفضاء

لنبدأ بشيء قريب من قلبي: استكشاف الفضاء. فقد مكنتنا تكنولوجيا الحواسيب من نقل الأفراد إلى القمر إرسال مركبة فضائية غير مأهولة لاستكشاف جميع الكواكب في نظامنا الشمسي، حتى الهبوط على قمر زحل تيتان وعلى مذنب. وكما سنستكشف في الفصل السادس، قد يساعدنا الذكاء الاصطناعي في المستقبل على استكشاف أنظمة ومجرات شمسية أخرى - إذا كان خالياً من أخطاء البرمجة Bugs. في الرابع من يونيو 1996 صاح العلماء -الذين كانوا يأملون بإجراء أبحاث حول الغلاف المغناطيسي للأرض- مبتهجين عندما أطلقت وكالةُ الفضاء الأوروبية ESA صاروخاً من طراز أريان 5 نحو السماء محملاً بالأدوات العلمية التي طوّروها، بعد 37 ثانية، اختفت ابتساماتهم

* إذا كنت ترغب في خريطة أكثر تفصيلاً من مشهد أبحاث سلامة الذكاء الاصطناعي، هناك خريطة تفاعلية نتيجة جهود مجتمعية بقيادة ريتشارد ملاح Richard Mallah من المعهد FLI: https://futureoflife.org/landscape

عندما انفجر الصاروخ في عرض ناري تكلفته مئات الملايين من الدولارات.[8] وكان السبب برمجية خاطئة تعالج أرقاماً أكبر بكثير من أن تستوعبه الـ 16 بتاً المخصصة لها.[9] وبعد مرور عامين دخلت مركبة كلايميت أوربيتور (مدار مناخ المريخ) Mars Climate Orbiter التابعة لناسا الغلاف الجوي للكوكب الأحمر وتعطلت عن العمل بسبب استخدام جزأين مختلفين من البرنامج لوحدات مختلفة لحساب القوة Force؛ مما تسبب في حدوث خطأ بنسبة 445% في التحكم في دفع الصاروخ.[10] وكان هذا خطأ ناسا الباهظ الثمن الثاني: إذ انفجرت مركبتهم مارينر 1 Mariner 1 المتجهة إلى كوكب الزهرة بعد إطلاقها من كيب كانافيرال Cape Canaveral في 22 يوليو 1962، بعد أن تعطل برنامج التحكم بسبب علامة ترقيم غير صحيحة.[11] ولإثبات أن الغربيين ليسوا وحدهم الذين أتقنوا فن إرسال أخطاء البرمجة إلى الفضاء؛ فقد فشلت مهمة فوبوس 1 Phobos 1 الروسية في 2 سبتمبر 1988، كانت هذه أثقل مركبة فضائية عابرة ما بين الكواكب يتم إطلاقها. وفشل الهدف المذهل المتمثل في إنزال مركبة الهبوط على سطح قمر المريخ فوبوس - عندما تسببت علامة فاصلة مفقودة في إرسال أمر "نهاية المهمة" End-of-mission إلى المركبة الفضائية حينما كانت في طريقها إلى المريخ؛ مما أدى إلى إغلاق جميع أنظمتها.[12]

ما نتعلمه من هذه الأمثلة هو أهمية ما يسميه علماء الحاسوب **التحقق** ***Verification***: ضمان أن البرنامج يُلبي جميع المتطلبات المتوقعة بالكامل. وكلما زادت الأرواح والموارد المعرضة للخطر، زادت الموثوقية من أن البرنامج يعمل على النحو المنشود. لحسن الحظ، قد يساعد الذكاء الاصطناعي على أتمتة وتحسين عملية التحقق. مثلاً، مؤخراً اختُبِرت نواة نظام تشغيل للأغراض العامة General-purpose operating-system تسمى ***seL4*** رياضياتيّاً مقابل مواصفات قياسية لتقديم ضمان قوي ضد الأعطال والعمليات غير الآمنة: على الرغم من أنها لا تأتي بعدُ مع أجراس وصفارات نظام تشغيل وندوز من ميكروسوفت Microsoft Windows وماك أو إس Mac OS، يمكنك أن تطمئن إلى أن نظام التشغيل هذا لن يؤدي إلى ما يعرف بـ«شاشة الموت الزرقاء» Blue screen of death أو "عجلة النهاية الدوّارة" Spinning wheel of doom.

لذا رعت وكالة مشاريع الأبحاث الدفاعية المتقدمة Defense Advanced Research Projects Agency (اختصاراً: الوكالة DARPA) تطويرَ مجموعة من أدوات الضمان العالي المفتوحة المصدر Open-source تسمى الأنظمة HACMS (الأنظمة العسكرية السيبرانية عالية الضمان High-assurance cyber military systems) والتي تُعدُّ آمنة تماماً. والتحدي المهم هو جعل هذه الأدوات قوية وسهلة الاستخدام بما فيه الكفاية بحيث يتم استخدامها على نطاق واسع. يتمثل التحدي الآخر في أن مهمة التحقق ذاتها ستزداد صعوبة مع انتقال البرمجيات إلى الروبوتات والبيئات الجديدة، ومع الاستعاضة البرمجيات التقليدية المبرمجة مسبقاً بأنظمة الذكاء الاصطناعي التي تستمر بالتعلم، ومن ثم تغيير سلوكها، كما في الفصل الثاني.

ذكاء اصطناعي للتمويل

يعد التمويل مجالاً آخر تغير بفعل تكنولوجيا المعلومات Information technology؛ مما سمح بإعادة تخصيص الموارد بشكل فعال في جميع أنحاء العالم بسرعة الضوء وتمكين التمويل بأسعار معقولة لكل شيء بدءاً من القروض العقارية إلى الشركات الناشئة Start-ups. ومن المحتمل أن يوفر التقدم في الذكاء الاصطناعي فرصاً رائعة مستقبلية من التداول المالي: فمعظم قرارات شراء/بيع سوق الأوراق المالية تُجرى الآن تلقائيّاً بواسطة أجهزة الحاسوب. وعادة ما يُغرَى طلبتي في المعهد MIT وطلبة الدراسات العليا في فريقي برواتب ابتدائية بأرقام فلكية ليعملوا على تحسين خوارزميات التداول.

فالتحقق خطوة مهمةٌ للبرمجيات المالية أيضاً. وقد تعلمت الشركة الأمريكية نايت كابيتال Knight Capital ذلك بالطريقة الصعبة في 1 أغسطس 2012. وذلك بفقدان 440 مليون دولار في غضون 40 دقيقة بعد تطبيق برنامج تداول لم يتم التحقق منه.[13] أما حادثة "الخسارة الساحقة" Flash crash بقيمة تريليون دولار في 6 مايو 2010؛ فقد كانت بسبب مختلف يستحق الالتفات إليه. فعلى الرغم من أنها تسببت في حدوث تقلبات هائلة لمدة نصف ساعة تقريباً قبل أن تستقر الأسواق، فقد تأرجحت فيه أسهم بعض الشركات البارزة مثل بروكتر وغامبل Procter & Gamble بين *سنت* واحد و100,000 دولار.[14] ولم تنجم المشكلة عن أخطاء برمجة Bug أو خلل في جهاز الحاسوب كان بالإمكان التحقق منها لتجنب ذلك. وبدلاً من ذلك، فقد كان سبب ذلك هو انتهاك التوقعات: وجدت برمجيات التداول التلقائي -المستخدمة من قبل العديد من الشركات- وجدت نفسها تعمل في وضع غير متوقع إذ كانت افتراضاتها غير صحيحة Valid - مثلاً، الافتراض بأنه إذا أبلغ حاسوب البورصة عن سعر السهم يعادل *سنتاً* واحداً، فإن سعر هذا السهم كان بالفعل يعادل *سنتاً* واحداً.

وحادثة الخسارة الساحقة توضح أهمية ما يسميه علماء الحاسوب **التدقيق** ***Validation***: بينما يسأل التحقّق «هل بنيت النظام كما يجب؟»، يسأل التدقيق هل بنيت النظام الصحيح؟»* مثلاً، هل يعتمد النظام على افتراضات ربما لا تكون صالحة دائماً؟ إذا كان الأمر كذلك، كيف يمكن تحسينه لمعالجة عدم اليقين بشكل أفضل؟

ذكاء اصطناعي للتصنيع

غني عن القول، إن الذكاء الاصطناعي يتضمن إمكانات كبيرة لتحسين التصنيع، من خلال التحكم في الروبوتات التي تعزز كلاً من الكفاءة والدقة. الآن، يمكن للطابعات

* بتعبير أدق، يسأل التحقق عما إذا كان النظام يفي بمواصفاته، في حين يسأل التدقيق عما إذا تم اختيار المواصفات الصحيحة.

ثلاثية الأبعاد التي تتطور باستمرار أن تصنع نماذج أولية لأي شيء بدءاً من مباني المكاتب إلى الأجهزة الميكانيكية الأصغر حجماً من حبة ملح.[15] وبينما تبني الروبوتات الصناعية الضخمة السياراتِ والطائرات، فإن المطاحن والألواح والقاطعات وما أشبه التي يتحكم فيها حاسوب ذو سعر في متناول اليد لم تعد تقتصر على إدارة المصانع، بل أيضاً متاحة لـ "حركة المصنّعين" Maker movement الشعبية، إذ يبني المتحمسون المحليون أفكارهم في أكثر من ألف مختبر من مختبرات فاب لاب Fab-Lab التي يديرها المجتمع في جميع أنحاء العالم.[16] ولكن كلما زادت الروبوتات حولنا، زادت أهمية أن نتحقق من برمجياتها وندققها. فأول شخص قُتِل بسبب روبوت هو روبرت ويليامز Robert Williams، وهو عامل في مصنع فورد في فلات روك، ميشيغان. وفي عام 1979، خرب الروبوت الذي كان من المفترض أن يحضر قطع الغيار من منطقة تخزين، فصعد روبرت إلى الطابق للوصول إلى الأجزاء بنفسه، غير أن الروبوت بدأ بهدوء بتحطيم رأسه لمدة 30 دقيقة إلى أن اكتشف زملاؤه في العمل ما قد حدث.[17] وضحية الروبوت التالي كان كينجي أورادا Kenji Urada، مهندس صيانة في مصنع كاواساكي في أكاشي باليابان، أثناء عمله على روبوت متعطل في عام 1981. وقام بطريق الخطأ بضغط مفتاح التشغيل فهصرته ذراع الروبوت الهيدروليكية حتى الموت.[18] وفي عام 2015 كان عامل يبلغ من العمر 22 عاماً في أحد مصانع إنتاج فولكس واغن في بوناتال بألمانيا، يعمل على تصنيع روبوت يقبض على قطع غيار السيارات ويحركها. ووقع خطأ ما؛ مما تسبب في إمساك الروبوت بالرجل وعصره في صفيحة معدنية ساحِقاً إياه حتى الموت.[19]



الشكل 3.3: بينما تكون الروبوتات الصناعية التقليدية باهظة الثمن وصعبة البرمجة، فإن هناك توجهاً نحو تطوير روبوتات تعتمد على الذكاء الاصطناعي الأقل تكلفة، قد تتعلم ماذا تفعل من العمال الذين ليس لديهم أي خبرة برمجة.

وعلى الرغم من أن هذه الحوادث مأساوية، فمن المهم الإشارة إلى أنها تشكل نسبة ضئيلة جدّاً من جميع الحوادث الصناعية. إضافة إلى ذلك، انخفضت الحوادث الصناعية بدلاً من ازديادها نظراً لتحسّن التكنولوجيا. إذ انخفضت من نحو 14000 حالة وفاة في عام 1970 إلى 4821 حالة وفاة في عام 2014 في الولايات المتحدة.[20] وتبين الحوادث الثلاثة المذكورة آنفاً أن إضافة الذكاء إلى الأجهزة الغبية بخلاف ذلك يجب أن تكون قادرة على زيادة تحسين السلامة الصناعية، من خلال جعل الروبوتات تتعلم أن تكون أكثر حذراً حول الناس. فالحوادث الثلاثة كان يمكن تجنبها مع التدقيق بشكل أفضل: تسببت الروبوتات في أضرار ليس بسبب الأخطاء أو الخبث، ولكن لأنها افترضت افتراضات غير صحيحة - أن الشخص لم يكن موجوداً أو أن الشخص كان قطعة غيار سيارة.

ذكاء اصطناعي للنقل

على الرغم من أن الذكاء الاصطناعي قد ينقذ حياة العديد من الأشخاص في التصنيع، فإنه قد ينقذ أكثر من ذلك في مجال النقل. تسببت حوادث السيارات وحدها في مقتل أكثر من 1.2 مليون شخص في عام 2015، كما تسببت حوادث الطائرات والقطارات والقوارب في مقتل الآلاف، في الولايات المتحدة. ومع وجود مقاييس السلامة العالية، فإن حوادث السيارات تسببت في مقتل نحو 35,000 شخص العام الماضي - سبعة أضعاف عدد الحوادث الصناعية مجتمعة.[21] وعندما ناقشنا هذا الموضوع في أوستن بتكساس في عام 2016 في الاجتماع السنوي لجمعية تقدم الذكاء الاصطناعي Association for the Advancement of Artificial Intelligence، جاشت مشاعر عالم الحاسوب موشيه فاردي حول هذا الموضوع وجادل في أنه ليس فقط بإمكان الذكاء الاصطناعي تقليل الوفيات الناجمة عن حوادث الطيران فحسب، بل يجب عليه ذلك: «إنها ضرورة أخلاقية!» كما قال، نظراً لحدوث جميع حوادث السيارات تقريباً بسبب خطأ بشري، يُعتقد على نطاق واسع أن السيارات ذاتية القيادة التي تعمل بنظام الذكاء الاصطناعي يمكنها القضاء على ما لا يقل عن 90% من الوفيات الناجمة عن حوادث الطرق، وهذا التفاؤل يغذي تقدماً كبيراً نحو وضع سيارات ذاتية القيادة فعليّاً على الطرق. إذ يتصور إيلون مسك Elon Musk أن السيارات ذاتية القيادة في المستقبل لن تكون أكثر أماناً فحسب، بل ستكسب أيضاً أموالاً لمالكيها عندما لا تكون هناك حاجة إليها، فتنافس أوبر Uber وليفت Lyft.

حتى الآن، تتمتع السيارات ذاتية القيادة بالفعل بسجل سلامة أفضل من السائقين البشر، والحوادث التي وقعت تؤكد على أهمية التدقيق وصعوبته. وحصل أول حادث مرور بسيط لسيارة ذاتية القيادة تابعة لشركة غوغل في 14 فبراير 2016؛ لأنها افترضت افتراضا خاطئاً بشأن الحافلة: أن سائقها سيفسح الطريق عندما تتجاوزه السيارة لتسير أمامه. أما أول حادث مميت؛ فقد وقع عندما اصطدمت سيارة تسلا ذاتية القيادة بمقطورة شاحنة تعبر الطريق السريع في 7 مايو 2016. ونجم ذلك عن افتراضين سيئين:[22] أن الجانب

الأبيض اللامع للمقطورة كان مجرد جزء من السماء الساطعة، وأن السائق (الذي يُقال إنه كان يشاهد فيلم «هاري بوتر» Harry Potter) سينتبه ويتدخل إذا حدث خطأ ما.*

لكن في بعض الأحيان، التحقق والتدقيق الجيدان ليسا كافيين لتجنب الحوادث؛ لأننا نحتاج أيضاً إلى **تحكم *Control*** جيد: قدرة المشغل البشري على مراقبة النظام وتغيير سلوكه أمر ضروري. ولكي تعمل مثل هذه الأنظمة التي تشمل **إنسان في الدائرة *Human-in-the-loop*** بشكل جيد، فمن المهم أن يكون الاتصال بين الإنسان والآلة فعالاً. وبهذه الروح، فإن اللون الأحمر على لوحة القيادة سينبهّك بسهولة إذا تركت باب صندوق سيارتك مفتوحة. وعلى النقيض من ذلك، عندما غادرت العبّارة البريطانية هيرالد أوف فري إنتربرايز Herald of Free Enterprise ميناء زيبروغ في 6 مارس 1987، وكان باب منصة التحميل مفتوحاً، لم يكن هناك ضوء تحذير أو أي تحذير مرئي آخر للقبطان، وانقلبت العبارة بعد وقت قصير من مغادرتها الميناء؛ مما أسفر عن مقتل 193 شخصاً.[23]

في ليلة 1 يونيو 2009 حدث عطل آخر مأساوي في التحكم كان من الممكن تجنبه من خلال تواصل أفضل بين الإنسان والآلة، عندما سقطت طيارة الخطوط الفرنسية رحلة رقم 447 في المحيط الأطلسي؛ مما أدى إلى مقتل 228 شخصاً كانوا على متنها. ووفقاً لتقرير صادر عن الحادث، «لم يفهم الطاقم قطّ أنهم عالقون في تيار الهواء، ومن ثم لم يطبقوا مطلقاً مناورة الانتعاش Recovery manoeuvre» - التي كانت ستشمل دفع خطم الطائرة إلى أسفل - حتى فات الأوان. وتكهن خبراء سلامة الطيران بأنه كان بالإمكان تجنّب الحادث لو كان هناك مؤشر "زاوية الهجوم" Angle-of-attack indicator في قمرة القيادة، ليدل الطيارين بأن خطم الطائرة ارتفع إلى الأعلى أكثر من اللازم.[24]

وعندما تحطمت طائرة خطوط أير إنتر Air Inter الرحلة 148 على جبال فوغس بالقرب من ستراسبورغ في فرنسا في 20 يناير 1992؛ مما أدى إلى مقتل 87 شخصاً، لم يكن السبب عدم وجود اتصال بين الإنسان والآلة، ولكن واجهة مستخدم User interface مثيرة للارتباك. فقد أدخل الطيارون العدد "33" على لوحة مفاتيح لأنهم أرادوا النزول بزاوية 3.3 درجة. ولكن الطيار الآلي فسر ذلك على أنه 3,300 قدم في الدقيقة لأنه كان في وضع Mode مختلف - وشاشة العرض كانت أصغر من أن تُظهِر الوضع بما كان سيسمح للطيارين بإدراك خطئِهم.

ذكاء اصطناعي للطاقة

تكنولوجيا المعلومات حققت العجائب بالنسبة إلى توليد الطاقة وتوزيعها، بالخوارزميات المتطورة التي توازن الإنتاج والاستهلاك عبر الشبكات الكهربائية في العالَم، وأنظمة

* حتى مع إدراج هذا العطل في الإحصائيات، وُجد أن القيادة الذاتية لسيارات تسلا تقلّل الحوادث بنسبة 40% عند تشغيل النظام: http://tinyurl.com/teslasafety.

التحكم المتطورة التي تُبقي محطات الطاقة تعمل بأمان وفعالية. ومن المحتمل أن يجعل مستقبل التقدم في الذكاء الاصطناعي "الشبكة الذكية" أكثر ذكاءً، للتأقلم على النحو الأمثل مع تغير العرض والطلب وصولاً إلى مستوى الألواح الشمسية المزدوجة على السطوح وأنظمة البطاريات المنزلية. ولكن في يوم الخميس الموافق 14 أغسطس عام 2003 انقطعت الكهرباء عن نحو 55 مليون شخص في الولايات المتحدة وكندا، وظل الكثير منهم من دون كهرباء لعدة أيام. هنا أيضاً، تم تحديد السبب الرئيسي بفشل الاتصالات بين الإنسان والآلة: منع خطأ برمجة Bug نظامَ الإنذار في غرفة التحكم في أوهايو من تنبيه المشغلين إلى الحاجة إلى إعادة توزيع الطاقة قبل حدوث مشكلة بسيطة (خطوط النقل المحملة بطاقة زائدة عن الحد ضربت أشجارا غير مُقلَّمة) تفاقمت حتى خرجت عن نطاق السيطرة.[25]

وفي 28 مارس 1979 أدى الانصهار النووي الجزئي في مفاعل بجزيرة ثري مايل في ولاية بنسلفانيا إلى نحو بليون دولار في تكلفة التنظيف وردة فعل كبير ضد الطاقة النووية. وحدد تقرير الحادث الأخير العديد من العوامل المساهمة في الحادث، بما في ذلك الارتباك الناجم عن ضعف واجهة المستخدم.[26] وعلى وجه الخصوص، ضوء التحذير الذي اعتقد المُشغِّلون أنه يشير إلى ما إذا كان صمام الأمان الحرج Safety-critical valve مفتوحاً أم مغلقاً كان يشير فقط إلى ما إذا كان قد تَمَّ إرسال إشارة لإغلاق الصمام - لذا لم يدرك المشغلون أن الصمام قد عَلِق وصار مفتوحاً.

حوادث الطاقة والنقل هذه تعلمنا أننا بينما نضع -بتزايد- الذكاء الاصطناعي في موضع المسؤولية عن الأنظمة المادية، فإننا نحتاج إلى بذل جهود بحثية جادة ليس فقط لجعل الآلات تعمل بشكل جيد من تلقاء نفسها، ولكن أيضاً لجعل الآلات تتعاون بفعالية مع وحدات التحكم البشرية. وعندما يصبح الذكاء الاصطناعي أكثر ذكاءً، فإن هذا لن يشمل فقط بناء واجهات مستخدم جيدة لتقاسم المعلومات، ولكن أيضاً تحديد كيفية تخصيص المهام على النحو الأمثل داخل فرق الإنسان والحاسوب - مثلاً، تحديد المواقف التي ينبغي نقل السيطرة فيها، ولتطبيق حكم الإنسان بكفاءة على القرارات ذات القيمة الأعلى بدلاً من تشتيت انتباه المراقبين مع سيل من المعلومات غير المهمة.

ذكاء اصطناعي للرعاية الصحية

للذكاء الاصطناعي إمكانات هائلة لتحسين الرعاية الصحية، فقد مكنت رقمنةُ السجلات الطبية الأطباءَ والمرضى من اتخاذ قرارات أسرع وأفضل، والحصول على مساعدة فورية من الخبراء في جميع أنحاء العالم في تشخيص الصور الرقمية. وفي الواقع، قريباً قد تكون أنظمة الذكاء الاصطناعي هي أفضل خبراء إجراء مثل هذا التشخيص، نظراً للتقدم السريع في رؤية الحاسوب Computer vision والتعلم

العميق Deep learning. مثلاً، أظهرت دراسة هولندية أجريت عام 2015 أن تشخيص الحاسوب لسرطان البروستاتا باستخدام التصوير بالرنين المغناطيسي (MRI) كان جيداً مثل تشخيص اختصاصي الأشعة البشري،[27] وأظهرت دراسة أجريت في ستانفورد عام 2016 أن الذكاء الاصطناعي قد يشخص سرطان الرئة باستخدام صور المجهر أفضل من اختصاصيي الأمراض Pathologists البشر.[28] وإذا كان تعلم الآلة قد يساعد على الكشف عن العلاقات بين الجينات والأمراض والاستجابة للعلاج، فإنه يمكن أن يحدث ثورة في الطب الشخصي، ويجعل حيوانات المزرعة أكثر صحة، ويُعزِّز من إنتاج محاصيل أكثر مرونة Resilient. إضافة إلى ذلك، فإن لدى الروبوتات القدرة على أن تكون جرّاحة أكثر دقة وموثوقية من البشر، حتى من دون استخدام الذكاء الاصطناعي المتقدم. فقد أُجريت مجموعة كبيرة ومتنوعة من العمليات الجراحية الآلية بشكل كامل في السنوات الأخيرة، وغالباً ما سمحت بالدقة والعمل على المستويات المتناهية الصغر، والشقوق الأصغر التي تؤدي إلى انخفاض فقدان الدم وتقليل الألم وتقليص وقت الشفاء.

للأسف، كانت هناك دروس مؤلمة حول أهمية البرمجيات القوية أيضاً في صناعة الرعاية الصحية، مثلاً، تم تصميم آلة العلاج بالإشعاع ثيراك 25 Therac-25 المصممة لعلاج مرضى السرطان لتعمل في وضعين مختلفين: إما مع شعاع منخفض الطاقة من الإلكترونات أو مع أشعة سينية عالية الطاقة Megavolt تُسلط على الهدف مع استخدام درع خاصة. ومع الأسف، أحياناً تسبب خطأ برمجة Bug لم يُتحقق منه في أن يقوم الفنيون بتسليط حزمة الموجات عالية الطاقة عندما اعتقدوا أنهم يسلطون حزمة منخفضة الطاقة، ومن دون استخدام الدرع مما أودى بحياة العديد من المرضى.[29] ومات العديدون أيضاً في المعهد الوطني للأورام National Oncologic Institute في بنما، إذ تمت برمجة معدات العلاج الإشعاعي التي تستخدم الكوبالت 60 المشع لأوقات التعرض المفرطة في عامي 2000 و 2001 بسبب واجهة مستخدم مربكة لم يتم التحقق من صحتها بشكل صحيح.[30] ووفقاً لتقرير حديث[31] حول الجراحة الروبوتية تم ربط بـ 144 حالة وفاة و1,391 حالة إصابة -في الولايات المتحدة بين عامي 2000 و2013 - بمشكلات شائعة لا تشمل فقط مشكلات الأجهزة مثل التقوّس الكهربائي Electrical arcing ونسيان قطع من الأدوات المحترقة أو المكسورة في المريض، ولكن أيضاً مشكلات البرمجيات مثل الحركات غير المنضبطة والتشغيل التلقائي.

والخبر السار هو أن ما يقرب من مليوني جراحة آلية مشمولة بالتقرير تمت بسلاسة، ويبدو أن الروبوتات تجعل الجراحة أكثر أماناً وليس أقل أماناً. فوفقاً لدراسة أجرتها حكومة الولايات المتحدة، تؤدي الرعاية السيئة في المستشفيات إلى أكثر من 100 ألف حالة وفاة سنويّاً بالولايات المتحدة وحدها.[32] ومن ثم، فإن الضرورة الأخلاقية لتطوير ذكاء اصطناعي أفضل للطب أكبر بكثير من ضرورة تطوير ذلك للسيارات ذاتية القيادة.

ذكاء اصطناعي للاتصالات

يمكن القول إن صناعة الاتصالات هي الصناعة التي كان للحواسيب أكبر تأثير فيها على الإطلاق حتى الآن، بعد تطوير لوحات مفاتيح الهاتف المحوسبة في خمسينات القرن العشرين، والإنترنت في ستيناته، وشبكة الإنترنت العالمية في عام 1989. ويستخدم الآن بلايين الأشخاص الإنترنت للتواصل أو التسوق أو قراءة الأخبار أو مشاهدة الأفلام أو لعب الألعاب، معتادين على امتلاك عالَم المعلومات بنقرة واحدة - وغالباً ما تكون مجاناً. ويعد **إنترنت الأشياء *Internet of things*** الناشئ بتحسن في الكفاءة والدقة والراحة والفوائد الاقتصادية من رفع كل شيء على الإنترنت من مصابيح وثيرموستات ومجمدات إلى شريحة مستجيب حيوية Biochip transponders مثبتة على حيوانات المزرعة.

وهذه النجاحات المذهلة في توصيل العالم ولدت تحدياً رابعاً لعلماء الحاسوب: فهم ليسوا بحاجة إلى تحسين التحقق والتدقيق والتحكم، ولكن أيضاً **الأمن *Security*** ضد البرمجيات الضارة ("البرمجيات الخبيثة" Malware) والقرصنة Hacks. وبينما نتجت جميع المشكلات المذكورة آنفاً من أخطاء غير مقصودة، فإن الأمن موجّه نحو **المخالفات المُتعمَّدة *Deliberate malfeasance***. وكانت أولى البرمجيات الخبيثة التي لفتت انتباه وسائل الإعلام إلى حد كبير هي ما يسمى دودة موريس Morris worm، التي أطلقت في 2 نوفمبر 1988، والتي استغلت أخطاء برمجة Bugs في نظام التشغيل يونيكس UNIX. ويُقال إنها كانت محاولة مضللة لحوسبة عدد أجهزة الحاسوب التي كانت متصلة بالإنترنت. وعلى الرغم من أنها أصابت وحطمت نحو 10% من أجهزة الحاسوب التي شكلت شبكة الإنترنت في ذلك الوقت، فإنها لم تحُل دون حصول منشِئُها، روبرت موريس، في النهاية على درجة أستاذ محاضر في علوم الحاسوب بالمعهد MIT.

البرمجيات الخبيثة الأخرى تستغل الثغرات الأمنية ليس في البرمجيات بل في الأفراد. في 5 مايو 2000، كما لو كانت للاحتفال بعيد ميلادي، تلقى العديد رسائل بريد إلكتروني مع سطر الموضوع "ILOVEYOU" من معارفهم و زملائهم ومن مستخدمي نظام تشغيل مايكروسوفت ويندوز Microsoft Windows الذين نقروا على المرفق «LOVE-LETTER-FOR-YOU،txt،vbs» فأطلقوا عن غير قصد برنامجاً نصيّاً أضر بحواسيبهم وأعاد إرسال البريد الإلكتروني إلى جميع من هم في دفتر عناوينهم. وتسببت هذه الدودة التي أنشأها اثنان من الشباب في الفلبين، في إصابة نحو 10% من الشبكة، تماماً كما فعلت دودة موريس، ولكن نظراً لأن الإنترنت كان أكبر بكثير في ذلك الوقت، فقد صارت تلك واحدة من أكبر الإصابات في جميع الأوقات. إذ أصابت أكثر من 50 مليون حاسوب وتسببت في أضرار مادية تفوق خمسة بلايين دولار. كما تعلمون -بألم- لا يزال الإنترنت ممتلئاً بأنواع لا حصر لها من البرمجيات الخبيثة المُعدية، التي يصنفها خبراء الأمن إلى ديدان Worms وأحصنة طروادة Trojans وفيروسات Viruses وغيرها من فئات السبر Sounding categories المرعبة. والأضرار التي تسببها تتراوح بين عرض رسائل

مزاح غير ضارة إلى مسح ملفاتك، أو سرقة معلوماتك الشخصية، أو التجسس عليك، أو اختطاف حاسوبك لإرسال رسائل غير مرغوب فيها Spam.

وبينما تستهدف البرمجيات الخبيثة أي حاسوب يمكنها استهدافه، يهاجم القراصنة Hackers أهدافاً محددة ذات أهمية - مثل أحدث الأمثلة البارزة محلات Target وتي جي ماكس TJ Maxx وأفلام سوني Sony Pictures وآشلي ماديسون Ashley Madison وشركة النفط السعودية أرامكو واللجنة الديمقراطية الوطنية الأمريكية U.S. Democratic National Committee. إضافة إلى ذلك، يبدو أن السرقات تزداد إثارة، فقد سرق المتسللون 130 مليون رقم بطاقة ائتمان ومعلومات أخرى عن الحوسبة من أنظمة هارتلاند للدفع Heartland Payment Systems في عام 2008. وفي عام 2013 انتهكوا أكثر من بليون (!) حساب بريد إلكتروني من ياهو.[33] وفي عام 2014 اخترق أحد موظفي إدارة شؤون الموظفين في حكومة الولايات المتحدة الأمريكية سجلات الموظفين ومعلومات طلبات التوظيف لأكثر من 21 مليون شخص، زُعم أن بينهم موظفين يحملون تصاريح أمنية عليا وبصمات أصابع الوكلاء السريين.

ونتيجة لذلك، أقلّب عيني كلما قرأت عن أن بعض الأنظمة الجديدة آمنة 100% وغير معرَّضة للتهديد. ولكن، "غير قابل للقرصنة" Unhackable هو بوضوح ما نحتاج إلى أن تكون عليه أنظمة الذكاء الاصطناعي في المستقبل قبل أن نسلمها زمام إدارة، مثلاً: البنية التحتية الحيوية أو أنظمة الأسلحة، ومن ثم فإن الدور المتنامي للذكاء الاصطناعي في المجتمع يواصل رفع مخاطر أمن الحاسوب. وبينما بعض القراصنة يستغل السذاجة البشرية أو النقاط الضعف المعقَّدة في البرمجيات التي تم إصدارها حديثاً، فإن البعض الآخر يتمكن من تسجيل الدخول غير المصرح به إلى الحواسيب البعيدة بالاستفادة من أخطاء برمجة Bugs بسيطة بقيت -وهو أمر محرج- دون أن يلاحظها أحد لفترة طويلة. واستمر خطأ البرمجة "هارتبليد" Heartbleed bug (نزيف القلب) من عام 2012 إلى عام 2014 في واحدة من أكثر مكتبات البرمجيات Software libraries شيوعاً للاتصال الآمن بين أجهزة الحاسوب، وتم تضمين البرمجة "باشدور" Bashdoor bug في نظام التشغيل ذاته لجميع أجهزة الحاسوب التي تعمل بالنظام يونكس Unix من 1989 حتى 2014. وهذا يعني أن أدوات الذكاء الاصطناعي لتحسين التحقق والتدقيق ستُحسن الأمن كذلك.

مع الأسف، يمكن أيضاً استخدام أنظمة الذكاء الاصطناعي الأفضل للعثور على ثغرات أمنية جديدة وتحقيق إنجازات خارقة أكثر تطوراً. تخيل، مثلاً، أن تصل إليك يوماً ما رسالة بريد إلكتروني "تصيد" Phishing مكتوبة بشكل شخصي تحاول إقناعك بالكشف عن معلومات شخصية، ترسل من حساب أحد أصدقائك من قِبل ذكاء اصطناعي اخترقه وانتحل شخصيته، وقلَّد أسلوبه في الكتابة بناءً على تحليل لرسائل البريد الإلكتروني الأخرى المرسلة، بما في ذلك الكثير من المعلومات الشخصية عنك من مصادر أخرى، هل ستصدق تلك الرسالة؟ ماذا لو ظهر أن رسالة البريد الإلكتروني الاحتيالي جاءت من شركة بطاقة الائتمان الخاصة بك وتلتها مكالمة هاتفية من صوت

بشري ودود لا يمكنك معرفة أنه تم إنشاؤه من قبل الذكاء الاصطناعي؟ في سباق التسلح المستمر لأمن الحاسوب بين الجريمة والدفاع، لا يوجد حتى الآن سوى القليل من المؤشرات على أن الدفاع يفوز.

قوانين

نحن البشر حيوانات اجتماعية أخضعت جميع الأنواع الأخرى واستحوذنا على الأرض بفضل قدرتنا على التعاون، وقد طورنا قوانين لتحفيز وتسهيل التعاون، لذلك إذا تمكن الذكاء الاصطناعي من تحسين أنظمتنا القانونية والحوكمية Governance؛ فسيمكننا ذلك من التعاون بشكل أكثر نجاحاً من أي وقت مضى؛ مما يبرز أفضل ما فينا. هناك الكثير من الفرص للتحسين هنا، سواء في كيفية تطبيق قوانيننا أم كيف كتابتها، لذلك دعونا نستكشف كلا منها بدوره.

ما أول ما يتبادر إلى ذهنك عندما تفكر في نظام المحاكم ببلدك؟ إذا كان هناك تأخير طويل وتكاليف مرتفعة وظلم في بعض الأحيان، فأنت لست وحدك. ألن يكون الأمر رائعاً إذا كانت خواطرك الأولى "الكفاءة" و"العدالة"؟ نظراً لأن العملية القانونية يمكن النظر إليها تجريدياً كعمليات حوسبية، تُدخل فيها معلومات حول الأدلة والقوانين ويكون المخرج هو القرار. ويحلم بعض العلماء بأتمتة العملية بالكامل باستخدام قضاة روبوتيين Robojudges: أنظمة ذكاء اصطناعي تطبق دون كلل المعايير القانونية العالية نفسها على كل حكم دون الخضوع لإنسان ودون أخطاء مثل التحيز والتعب أو عدم اطلاع على أحدث معرفة.

قضاة روبوتيون

في عام 1963 أُدين بايرون دي لا بيكويث جونيور De La Beckwith Jr، في عام 1994 باغتيال زعيم الحقوق المدنية مدغار إيفرز Medgar Evers. ولكن محكمتين منفصلتين في الميسيسيبي متألفتين من البيض كانتا قد فشلتا في إدانته في العام الذي تلا القتل مباشرة، على الرغم من أن الأدلة المادية كانت هي نفسها في الأساس.[34] للأسف، فإن التاريخ القانوني ممتلىء بالأحكام المنحازة بسبب لون البشرة والجنس والتوجه الجنسي والدين والجنسية وعوامل أخرى. وقد يضمن القضاة الروبوتيون، من حيث المبدأ، لأول مرة في التاريخ، أن الجميع متساوون حقّاً بموجب القانون: ويمكن برمجتهم ليكونوا متطابقين ويعاملون الجميع على قدم المساواة، وعلى تطبيق القانون بشفافية بطريقة غير منحازة حقاً.

القضاة الروبوتيون يمكنهم أيضاً القضاء على التحيزات البشرية التي هي عرضية وليست متعمدة. مثلاً، زعمت دراسة مثيرة للجدل في عام 2012 لقضاة أنهم أصدروا

أحكاماً أشد صرامة عندما كانوا جوعى: فبينما رفضوا الموافقة على نحو 35% من حالات الإفراج المشروط بعد الإفطار مباشرة، فإنهم رفضوا أكثر من 85% قبل الغداء مباشرة.[35] والقصور الآخر في القضاة البشر هو أنهم قد يفتقرون إلى الوقت الكافي لاستكشاف جميع تفاصيل القضية. وفي المقابل، يمكن نسخ ملفات القضاة الروبوتيين بسهولة، لأنها تتكون مما هو أكثر قليلاً من البرمجيات؛ مما يسمح بمعالجة جميع الحالات المعلقة بالتوازي بدلاً من على التوالي، إذ تحصل كل حالة على ملف القاضي الروبوتي الخاص بها. أخيراً، على الرغم من أنه من المستحيل على القضاة البشر إتقان كل المعرفة التقنية اللازمة لكل حالة ممكنة، من نزاعات براءات الاختراع الشائكة إلى أسرار القتل التي تتوقف على أحدث العلوم الجنائية، فقد تكون لدى القاضي الروبوتي في المستقبل ذاكرة غير محدودة وقدرة تعلم غير محدودة.

وفي يوم من الأيام قد يكون مثل هؤلاء القضاة الروبوتيين أكثر فاعلية وعدلاً، بحكم كونهم غير منحازين وذوي كفاءة وشفافية، إن كفاءتهم تجعلهم أكثر عدالة: ومن خلال تسريع العملية القانونية يصبح من الصعب على المحامين الأذكياء تشويه النتيجة. كما قد يجعل القضاة الروبوتيون الحصول على العدالة من خلال المحاكم أرخص بشكل كبير. وهذا قد يزيد بشكل كبير من فرص الفرد أو الشركة الناشئة التي تعاني ضائقة مالية أن تكسب قضية ضد بليونير أو شركة متعددة الجنسيات ذات جيش من المحامين.

من ناحية أخرى، ماذا لو كان لدى القضاة الروبوتيون أخطاء برمجة Bugs أو تعرضوا للاختراق؟ كلتا الحالتين قد أثرتا بالفعل في عمليات التصويت الآلي. وعندما تكون سنوات وراء القضبان أو الملايين في البنك هي التي على المحك، تكون الحوافز للهجمات الإلكترونية أكبر. وحتى إذا كان من الممكن جعل الذكاء الاصطناعي قوياً بدرجة كافية حتى يتسنى لنا أن نثق بأن القاضي الروبوتي يستخدم القانون الشرعي، فهل سيشعر الجميع بأنهم يفهمون أسبابه المنطقية بدرجة كافية لاحترام حكمه؟ يتفاقم هذا التحدي بسبب النجاح الأخير الذي حققته الشبكات العصبية، التي غالباً ما تتفوق على خوارزميات الذكاء الاصطناعي التقليدية، لكن يكون الثمن هو عدم وضوح كيفية توصلها إلى النتيجة. وإذا رغب المدعى عليهم في معرفة سبب إدانتهم، ألا ينبغي أن يكون لهم الحق في الحصول على إجابة أفضل من: «لقد دربنا النظام على الكثير من البيانات، وهذا ما قرر»؟ إضافة إلى ذلك، أظهرت الدراسات الحديثة أنه إذا درّبت نظاماً عصبياً عميقاً يحتوي على كميات هائلة من بيانات السجناء، فيمكنه التنبؤ بمن يحتمل أن يعود إلى الجريمة (ومن ثم يجب حرمانه من الإفراج المشروط) أفضل من القضاة البشر. ولكن ماذا لو وجد هذا النظام - أن احتمال العودة إلى حياة الجريمة مرتبط إحصائياً بجنس السجين أو عرقه - هل يعتبر هذا بمثابة تحيز جنسي أو عنصري يحتاج إلى إعادة البرمجة؟ في الواقع، جادلت دراسة أجريت عام 2016 في أن برمجيات التنبؤ بالعودة إلى الجريمة المستخدمة في جميع أنحاء الولايات المتحدة كانت منحازة ضد الأمريكيين الإفريقيين وساهمت في إصدار أحكام غير عادلة.[36] وهذه أسئلة مهمة نحتاج جميعاً

إلى التفكير فيها ومناقشتها لضمان بقاء الذكاء الاصطناعي مفيداً. نحن لا نواجه قراراً من نوع إما كل الأمر أو لا شيء منه بخصوص القضاة الروبوتيين، بل قراراً بشأن مدى وسرعة تطبيق الذكاء الاصطناعي في نظامنا القانوني. هل نريد أن تكون لدى القضاة البشر أنظمة لدعم اتخاذ القرار تعتمد على الذكاء الاصطناعي، مثلهم مثل أطباء الغد؟ هل نريد أن نذهب إلى أبعد من ذلك ونتخذ قرارات بشأن الحكم الذي يمكن استئنافه أمام قضاة من البشر، أو هل نريد أن نمضي إلى أبعد حدٍّ ونعطي الكلمة النهائية للأجهزة، حتى بالنسبة إلى عقوبات الإعدام؟

جدال قانوني

حتى الآن، استكشفنا **تطبيق** القانون فقط؛ دعنا الآن ننتقل إلى محتواه. فهناك إجماع واسع على أن قوانيننا بحاجة إلى التطور لمواكبة تقنيتنا. مثلاً، تمت تبرئة اثنين من المبرمجين الذين أنشؤوا دودة ILOVEYOU المذكورة آنفا وتسببوا في أضرار ببلايين الدولارات وسرِّحوا دون أي غرامة؛ لأنه في ذلك الوقت لم تكن هناك قوانين ضد إنشاء البرمجيات الخبيثة في الفلبين. ولما كانت وتيرة التقدم التكنولوجي تتسارع، فيجب تحديث القوانين بسرعة أكبر، لأنها تميل إلى التأخر عن مواكبة التطورات. ربما يكون إدخال مزيد من اختصاصيي التكنولوجيا المهرة في كليات الحقوق والحكومات خطوة ذكية للمجتمع. لكن هل يجب أن يتبع ذلك دعم نظم القرار القائم على الذكاء الاصطناعي للناخبين والمشرعين، ويتلوه مشرعون روبوتيون بالكامل؟

إن أفضل طريقة لتغيير قوانيننا لتعكس التقدم في الذكاء الاصطناعي هي موضوع مثير للجدل لدرجة آسرة للخيال. وأحد الخلافات يعكس التوتر بين الخصوصية مقابل حرية المعلومات، يجادل محبو الحرية في أنه كلما قلت *خصوصيتنا*، زاد عدد الأدلة المتاحة للمحاكم، وكانت الأحكام أكثر عدلاً. مثلاً، إذا دخلت الحكومة في الأجهزة الإلكترونية للجميع لتسجيل مكان وجودهم وما يكتبونه، وما ينقرون عليه، وما يقولونه وما يفعلونه على الإنترنت، فسيتم حل العديد من الجرائم بسهولة، ويمكن منع وقوع جرائم إضافية. ويعارض دعاة الخصوصية ذلك لأنهم لا يريدون دولة مراقبة أورويلية Orwellian، وحتى إذا فعلوا ذلك، فهناك خطر من تحولها إلى ديكتاتورية شمولية ذات أبعاد ملحمية. إضافة إلى ذلك، فقد تحسنت تقنيات تعلم الآلة في تحليل بيانات المخ من ماسحات التصوير بالرنين المغناطيسي الوظيفي fMRI لتحديد ما يفكر فيه الشخص، وعلى وجه الخصوص، ما إذا كان يقول الحقيقة أو الكذب.[37] وإذا صارت تقنية مسح الدماغ بمساعدة الذكاء الاصطناعي شائعة في قاعات المحاكم، فقد تكون العملية الشاقة الحالية المتمثلة في إثبات وقائع القضية بسيطة ومبسطة بشكل كبير؛ مما يتيح إجراء محاكمات أسرع وإصدار أحكام أكثر عدالة. ولكن دعاة الخصوصية قلقون من أن مثل هذه الأنظمة ترتكب أخطاء في بعض الأحيان. والأهم من ذلك، يتساءلون

عمّا إذا كانت عقولنا يجب أن تكون محظورة على التطفل الحكومي. فالحكومات التي لا تدعم حرية الفكر قد تستخدم هذه التكنولوجيا لتجريم اعتناق بعض المعتقدات والآراء. أين ترسم -أنت- الخط الفاصل بين العدالة والخصوصية، وبين حماية المجتمع وحماية الحرية الشخصية؟ أينما ترسمه، فهل يتحرك تدريجيّاً ولكن بلا هوادة نحو تقليل الخصوصية للتعويض عن حقيقة أن الأدلة تزداد سهولة تزويرها؟ مثلاً، بمجرد أن يتمكن الذكاء الاصطناعي من إنشاء مقاطع فيديو وهمية واقعية بالكامل ترتكب فيها جرائم، هل ستصوت لصالح نظام تتعقب فيه الحكومة مكان كل فرد في جميع الأوقات، وتزودك بشاهد إثبات غير قابل للنقض إذا لزم الأمر؟

تناقض آسر آخر هو ما إذا كان ينبغي تنظيم أبحاث الذكاء الاصطناعي أو، بشكل أعم، أي حافز ينبغي لصانعي السياسات منحه للباحثين في الذكاء الاصطناعي لتحقيق أقصى قدر ممكن من النتائج المفيدة. وجادل بعض الباحثين في الذكاء الاصطناعي ضد جميع أشكال تنظيم تطوير الذكاء الاصطناعي، بدعوى أنهم سيؤخرون دون داعٍ الابتكار المطلوب بشكل عاجل (مثلاً، السيارات الذاتية القيادة المنقذة للحياة)، وسيدفعون أبحاث الذكاء الاصطناعي المتطورة إلى الاختباء و/أو الانتقال إلى بلدان أخرى ذات حكومات أكثر تساهلاً. وفي مؤتمر بورتوريكو للذكاء الاصطناعي المفيد المذكور في الفصل الأول، جادل إيلون ماسك في أن ما نحتاج إليه الآن من الحكومات ليس رقابة بل بصيرة: على وجه التحديد، أشخاص لديهم القدرة التقنية في مناصب حكومية يستطيعون مراقبة تقدم الذكاء الاصطناعي وتوجيهه إذا حاد عن الطريق. كما جادل في أن التنظيم الحكومي قد يعزز أحياناً التقدم بدلاً من تأخيره: مثلاً، إذا كانت معايير السلامة الحكومية للسيارات ذاتية القيادة قد تساعد على تقليل عدد حوادث السيارات ذاتية القيادة، فإن ردة الفعل العام الرافضة تكون أقل احتمالاً ويمكن تسريع تبني التكنولوجيا الجديدة. لذا، قد تفضل معظم شركات الذكاء الاصطناعي الواعية بالسلامة التنظيمَ الذي يجبر المنافسين الأقل صرامة على مطابقة معايير السلامة العالية.

وهناك جدل قانوني آخر مثير للاهتمام حول منح الآلات حقوقاً، إذا قلصت السيارات ذاتية القيادة عدد الوفيات السنوية التي بلغت 32,000 شخص في الولايات المتحدة إلى النصف، فربما لن تحصل شركات صناعة السيارات على 16,000 رسالة شكر، بل 16,000 دعوى قضائية. لكن إذا تسببت سيارة ذاتية القيادة في وقوع حادث، فمن يجب أن يكون مسؤولاً - شاغلها أو مالكها أو الشركة المصنعة لها؟ اقترح الباحث القانوني ديفيد فلاديك David Vladeck إجابة رابعة: السيارة نفسها! على وجه التحديد، يقترح السماح للسيارات ذاتية القيادة باستخراج تأمين على السيارات (بل يُتطلَّب منها ذلك)، وبهذه الطريقة، تتأهل الطرز التي لها سجل أمان ممتاز للحصول على تأمين منخفض جدّاً، وربما تكون أدنى مما هو متاح للسائقين من البشر. وبينما ستؤهل فقط النماذج متدنية الضعف من الشركات المصنعة والمنخفضة الجودة للحصول على بوالص التأمين، فإن ذلك سيجعلها باهظة التكاليف لامتلاكها.

ولكن إذا سمح لآلات مثل السيارات بالحصول على بوالص تأمين، فهل يجب عليها أيضاً امتلاك الأموال والممتلكات؟ إذا كان الأمر كذلك، فليس هناك ما يمنع الحواسيب الذكية قانوناً من جني الأموال من سوق الأوراق المالية واستخدامها لشراء الخدمات عبر الإنترنت. وبمجرد أن يبدأ الحاسوب بدفع أجور البشر للعمل من أجله، فقد يحقق أي شيء يمكن للبشر فعله. وإذا صارت أنظمة الذكاء الاصطناعي في النهاية أفضل من البشر في الاستثمار (وهي تحقق ذلك بالفعل في بعض المجالات)، فقد يؤدي ذلك إلى وضع يكون فيه معظم اقتصادنا مملوكاً ومسيطراً عليه بواسطة الآلات، هل هذا ما نريد؟ إذا بدا هذا بعيداً، خذ بعين الاعتبار أن معظم اقتصادنا مملوك بالفعل لشكل آخر من أشكال الكيانات غير البشرية: الشركات التي غالباً ما تكون أقوى من أي شخص فيها ويمكنها إلى حد ما أن تكون لها حياة خاصة بها.

إذا كنت موافقاً على منح الآلات حقوق التملك، فما رأيك في منحهم حق التصويت؟ إذا كان الأمر كذلك، فهل يجب أن يحصل كل برنامج حاسوبي على صوت واحد، على الرغم من أنه بإمكانه إنتاج تريليونات النسخ من نفسه في السحابة إذا كانت غنية بما يكفي؛ مما يضمن أن البرنامج الحاسوبي سيقرر نتيجة الانتخابات؟ إذا لم يكن الأمر كذلك، فما هو الأساس الأخلاقي الذي نميز بناء عليه ضد العقول الآلية مقارنة بالعقول البشرية؟ هل يحدث فرقاً إذا كانت عقول الآلات واعية بمعنى أن تكون لديها تجربة ذاتية كما نفعل؟ سنستكشف هذه الأسئلة المثيرة للجدل المتعلقة بتحكم الحاسوب في عالمنا بتعمق أكبر في الفصل التالي، والأسئلة المتعلقة بوعي الآلة في الفصل الثامن.

أسلحة

منذ أزمان سحيقة والبشرية تعاني المجاعة والمرض والحرب، فقد ذكرنا بالفعل الكيفية التي يمكن بها أن يساعد الذكاء الاصطناعي على الحد من المجاعة والمرض، فماذا عن الحرب؟ يجادل البعض في أن الأسلحة النووية تردع الحرب بين الدول التي تمتلكها لأنها مرعبة جدّاً، فماذا عن السماح لجميع الدول بأن تبني ترسانة أسلحة تقوم على الذكاء الاصطناعي لإنهاء الحرب ككل إلى الأبد؟ إذا كنت غير مقتنع بهذه الحجة وتعتقد أن حروب المستقبل أمر لا مفر منه، فما رأيك في استخدام الذكاء الاصطناعي لجعل هذه الحروب إنسانية؟ إذا كانت الحروب تتكون فقط من آلات، فإن الآلات لا تحتاج إلى قتل جنود أو مدنيين. إضافة إلى ذلك، نأمل بأن تكون الطائرات التي تعمل من دون طيار في المستقبل وغيرها من أنظمة الأسلحة الذاتية Autonomous weapon systems (اختصاراً: الأسلحة AWS)، المعروفة أيضاً من قبل خصومها بـ "الروبوت القاتل" Killer robots أكثر عدلاً وعقلانية من الجنود البشر: مجهزة بمجسات فائقة ولا تخشى التعرض للقتل، قد تظل هادئة، وتتمالك أعصابها وتحسب الأمور حتى في خضم المعركة، فتكون أقل عرضة لقتل المدنيين بطريق الخطأ.



الشكل 4.3: في حين أن الطائرات العسكرية من دون طيار اليوم (مثل طائرة سلاح الجو الأمريكي MQ- 1 Predator) يتم التحكم فيها عن بعد من قبل البشر، فإن لدى الطائرات من دون طيار التي تعمل بالذكاء الاصطناعي في المستقبل القدرة على إخراج البشر من دائرة اتخاذ القرار واستخدام خوارزمية لتحديد من تستهدف وتقتل.

إنسان في الدائرة

لكن ماذا لو كانت الأنظمة الآلية ممتلئة بأخطاء برمجة Bugs، مرتبكة أو لا تتصرف كما هو متوقع؟ نظام الكتائب Phalanx system الأمريكي للطرادات من طراز إيجيس Aegis-class cruisers يكشف عن التهديدات والهجمات مثل الصواريخ المضادة للصواريخ والطائرات. مثلاً كانت السفينة **يو. إس. إس. فِنسنس** ***USS Vincennes*** طراداً بصواريخ موجهة يطلق عليها تحببا اسم روبوكروزر Robocruiser في إشارة إلى نظام إيجيس Aegis. وفي 3 يوليو 1988، في خضم مناوشات مع الزوارق الحربية الإيرانية خلال الحرب الإيرانية العراقية، حذر نظام الرادار من وجود طائرة قادمة. واستنتج النقيب ويليام رودجرز الثالث Captain William Rodgers III أنهم يتعرضون للهجوم من قبل طائرة مقاتلة إيرانية من طراز F-14 تنقض عليهم. وأعطى النقيب نظامَ إيجيس الموافقة على إطلاق النار. وما لم يدركه في ذلك الوقت هو أنهم أسقطوا طائرة الخطوط الإيرانية رحلة رقم 655، وهي طائرة ركاب إيرانية مدنية؛ مما أسفر عن مقتل 290 شخصاً على متنها والتسبب في غضب دولي. وأشار التحقيق اللاحق إلى وجود واجهة مستخدم مربكة لم تُظهر تلقائيّاً أياً من النقاط على شاشة الرادار هي طائرات مدنية (اتبعت الرحلة رقم 655 مسارها اليومي المعتاد، وكان جهاز متلقي إشارات الطيارات المدنية Civilian aircraft transponder يعمل) أو النقاط التي كانت تنازلية (بالنسبة إلى الهجوم) أو أي نقاط كانت تهبط (كما في حالة الهجوم) أو تقلع (كما كانت الرحلة 655 بعد الإقلاع من طهران). وبدلاً من ذلك، عندما تم استعلام النظام الأوتوماتيكي للحصول على معلومات حول الطائرة

الغامضة، فقد أبلغ النظام عن "هبوط" لأن ذلك كان حالة طائرة مختلفة أعاد النظام تعيينها بشكل مرتبك باستخدام عدد تستخدمه البحرية لتتبع الطائرات: ما كان يهبط -بدلاً من ذلك- كانت طائرة دورية جوية قتالية أمريكية Surface combat air patrol plane تعمل بعيداً في بحر عمان.

وفي هذا المثال، كان هناك إنسان في الدائرة يتخذ القرار النهائي، والذي تحت وطأة الضغط وثق كثيرا فيما أخبره به النظام الآلي. حتى الآن، وفقاً لمسؤولي الدفاع حول العالم، فإن جميع أنظمة الأسلحة تشتمل على إنسان في حلقة اتخاذ القرار، باستثناء المفخّخات Booby traps ذات التكنولوجيا المنخفضة مثل الألغام الأرضية. ولكن، يجري حالياً تطوير أسلحة مستقلة حقّاً تقوم باختيار ومهاجمة الأهداف من تلقاء نفسها. ومن المغري عسكريّاً إخراج جميع البشر من دائرة القرار لصالح اكتساب السرعة: في قتال بين طائرة من دون طيار تتمتع بالسيطرة الذاتية الكاملة يمكنها الاستجابة على الفور وطائرة من دون طيار تتفاعل بشكل أكثر بطئاً؛ لأن إنساناً يتحكم فيها عن بعد عبر نصف العالم، هل تعتقد أنها ستفوز؟

ولكن، كانت ستقع حوادث وشيكة لولا أننا كنا محظوظين جدّاً لوجود إنسان في الدائرة. وفي 27 أكتوبر 1962، أثناء أزمة الصواريخ الكوبية، حاصرت 11 مدمرة تابعة للبحرية الأمريكية وحاملة الطائرات يو. إس. إس. **راندولف** USS ***Randolph*** الغواصة السوفييتية B-59 بالقرب من كوبا، في المياه الدولية خارج منطقة "الحَجْر" الأمريكية. وما لم يعرفوه هو أن درجة الحرارة على متن الغواصة ارتفعت فوق 45 درجة سيليزية (113 درجة فهرنهايت) لأن بطاريات الغواصة نفدت وتكييف الهواء توقف، وكان الطاقم على وشك التسمم بثاني أكسيد الكربون. وكان العديد من أفراد الطاقم قد أغمي عليهم، ولم يكن الطاقم على اتصال بموسكو لعدة أيام، ولم يكن يعلم ما إذا كانت الحرب العالمية الثالثة قد بدأت بالفعل. ثم بدأ الأمريكيون بإسقاط الشحنات العميقة الصغيرة التي أخبروا موسكو، دون علم الطاقم، بأنها تهدف فقط إلى إجبار الغواصة إلى الصعود إلى السطح والمغادرة. «لقد ظننا أن هذه هي النهاية»، كما يتذكر عضو الطاقم في. بي. أورلوف V. P. Orlov، ويقول: «شعرت كما لو كنت جالساً في برميل معدني، يهاجمه شخص باستمرار بمطرقة ثقيلة». وما لم يعرفه الأمريكيون أيضاً هو أن طاقم الغواصة B-59 كان لديه طوربيد نووي، يُؤذَن لإطلاقه من دون الرجوع إلى موسكو، بالطبع.

قرر الكابتن سافيتسكي Captain Savitski إطلاق الطوربيد النووي، وتعجب فالنتين غريغوريفيتش Valentin Grigorievich ضابط إطلاق الطوربيد قائلاً: «سنموت. لكننا سنغرقهم جميعاً - لن نخزي قوتنا البحرية!». ولحسن الحظ كان قرار إطلاق الطوربيد يحتاج إلى موافقة ثلاثة ضباط على ظهر الغواصة، وأحدهم فاسيلي أركيبوف Vasili Arkhipov، الذي قال لا. ومن المحزن أن قلة قليلة قد سمعت عن أركيبوف، على الرغم من أن قراره ربما جنبنا حرباً عالمية ثالثة وكان ذلك الحدثُ الوحيد الإسهامَ الأكثرَ قيمةً للإنسانية في التاريخ الحديث.[38] ومن المعقول أيضاً التفكير فيما قد يحدث لو كانت الغواصة B-59 مستقلة ذاتيّاً يحكمها ذكاء اصطناعي من دون وجود بشر في الدائرة.

بعد عقدين من الزمن، في 9 سبتمبر 1983، اشتد التوتر مرة أخرى بين القوتين العظميين: كان الرئيس الأمريكي رونالد ريغان Ronald Reagan قد وصف الاتحاد السوفييتي مؤخراً بأنه "إمبراطورية شريرة" Evil empire. وفي الأسبوع السابق كان الاتحاد السوفييتي قد أسقط طائرة ركاب تابعة لشركة الخطوط الجوية الكورية ضلت طريقها إلى مجاله الجوي؛ مما أسفر عن مقتل 269 شخصاً - بمن فيهم مستشار أمريكي. بعد ذلك أعلن نظام الإنذار المبكر السوفييتي الآلي أن الولايات المتحدة أطلقت صواريخ نووية برية على الاتحاد السوفييتي، تاركة لسيستانيسلاف بيتروف Stanislav Petrov مجرد دقائق ليقرر ما إذا كان هذا إنذاراً خاطئاً. وكان القمر الاصطناعي يعمل بشكل صحيح، ومن ثم فإن البروتوكول التالي يدفعه إلى الإبلاغ عن هجوم نووي وارد. وبدلاً من ذلك، فقد وثق في غريزته، إذ أكد أنه من غير المرجح أن تهاجم الولايات المتحدة بصواريخ فقط، وأبلغ قادته أنه كان إنذاراً خاطئاً دون أن يعلم أن هذا صحيح. واتضح فيما بعد أن أقماراً اصطناعية أخطأت في فهم انعكاسات الشمس من قمم السحب على أنها شعلات نيران المحركات الصاروخية.[39] أتساءل ماذا كان سيحدث لو تمت الاستعاضة عن بيتروف بنظام ذكاء اصطناعي اتبع بروتوكولاً حرفيّاً.

سباق الأسلحة التالي؟

كما تكون قد خمَّنت بلا شك الآن، أنا شخصيّاً لدي مخاوف جادة حول أنظمة الأسلحة الذاتية. ولكنني لم أبدأ حتى الآن بإخبارك عن قلقي الرئيسي: نقطة نهاية سباق تسلح أسلحة الذكاء الاصطناعي. وفي يوليو 2015 أعربت عن هذا القلق في الرسالة المفتوحة التالية مع ستيوارت راسل، مع ملاحظات مفيدة من زملائي في معهد مستقبل الحياة Future of Life Institute:[40]

الأسلحة الذاتية:

خطاب مفتوح من الباحثين في الذكاء الاصطناعي والروبوتات

> تحدِّد الأسلحة الذاتية التحكم الأهداف وتصوبها دون تدخل بشري، مثلاً، تشتمل هذه المجموعات على كوادكوبترات (طائرات مروحية صغيرة متعددة المراوح) Quadcopters مسلحة يمكنها البحث عن الأشخاص الذين يستوفون معايير محددة معينة وتصفيتهم، ولكنها لا تشمل صواريخ كروز أو طائرات من دون طيار موجهة عن بعد يتخذ البشر من خلالها جميع قرارات الاستهداف. ووصلت تقنية الذكاء الاصطناعي إلى نقطة سيكون فيها استخدام مثل هذه الأنظمة عمليّاً إن لم يكن ممكناً من الناحية القانونية

في غضون سنوات وليس عقود. والمخاطر كبيرة: وصفت الأسلحة الذاتية بأنها الثورة الثالثة في الحرب، بعد البارود والسلاح النووي.

لقد قُدمت حجج كثيرة سواء لصالح الأسلحة الذاتية أو ضدها. مثلاً: إن الاستعاضة عن الجنود البشر بالآلات تعد أمراً جيداً من حيث تقليل الإصابات للمالك، ولكنها سيئة لأنها تخفِّض عتبة نشوب حرب. والسؤال الرئيسي للبشرية اليوم هو ما إذا كان يجب بدء سباق تسلح ذكاء اصطناعي عالمي أو حظره من البداية. إذا استمرت أي قوة عسكرية كبرى في تطوير أسلحة الذكاء الاصطناعي، فإن بدء سباق التسلح العالمي سيكون أمراً واقعا لامحالة تقريباً، ونقطة النهاية لهذا المسار التكنولوجي واضحة: الأسلحة الذاتية ستكون كلاشينكوف الغد. وعلى عكس الأسلحة النووية، فإنها لا تحتاج إلى مواد خام مكلفة أو يصعب الحصول عليها، لذلك ستصبح متاحة في كل مكان ورخيصة بالنسبة إلى جميع القوى العسكرية الكبيرة لإنتاجها على نطاق واسع. وسيكون الأمر مسألة وقت فقط حتى تظهر في السوق السوداء وفي أيدي إرهابيين وديكتاتوريين يرغبون في التحكم تحكماً أكبر في شعوبهم، وأمراء الحرب Warlords الذين يسعون إلى التطهير العرقي، وما إلى ذلك. والأسلحة الذاتية مثاليةٌ لمهام مثل الاغتيالات، وزعزعة استقرار الدول، وتقسيم الشعوب والقتل الانتقائي لجماعة عرقية معينة، لذلك نعتقد أن سباق تسلح الذكاء الاصطناعي العسكري لن يكون مفيداً للبشرية. ولكن هناك العديد من الطرق التي يمكن بها للذكاء الاصطناعي جعل حقول المعركة أكثر أماناً للبشر، وخاصة المدنيين، وذلك من دون تطوير أدوات جديدة لقتل الأفراد.

ومثلما لا يعبأُ معظم الكيميائيين والعلماء بصنع أسلحة كيميائية أو بيولوجية، فإن معظم الباحثين في الذكاء الاصطناعي ليسوا مهتمين بتطوير أسلحة الذكاء الاصطناعي ولا يريدون للآخرين تشويه مجالهم من خلال القيام بذلك. لأنه من المحتمل أن يؤدي ذلك إلى ردة فعل عامة كبيرة ضد الذكاء الاصطناعي مما يحد من فوائده الاجتماعية المستقبلية. في الواقع، أيد كيميائيون وبيولوجيون على نطاق واسع الاتفاقيات الدولية التي نجحت في حظر الأسلحة الكيميائية والبيولوجية، كما أيد معظم الفيزيائيين معاهدات حظر الأسلحة النووية وأسلحة الليزر الموجّهة من الفضاء والمسببة للعمى.

ولحماية قلقنا من أن يُهمَّش على أنه مجرد مخاوف صادرة عن محبي أشجار مسالمين، أردت أن يوقع على خطابنا المفتوح أكبر عدد ممكن من الباحثين المتخصصين بالذكاء الاصطناعي وعلماء الروبوتات. وكانت الحملة الدولية للحد من الأسلحة الذاتية قد جمعت

من قبل المئات من الموقعين الذين طالبوا بفرض حظر على الروبوتات القاتلة، وكنت أظن أنه يمكننا القيام بعمل أفضل، وكنت أعلم أن المنظمات المهنية ستكون مترددة في مشاركتنا قوائمهم من عناوين البريد الإلكتروني الضخمة الخاصة بالأعضاء لغرض يمكن تفسيره على أنه سياسي، لذا جمعت قوائم بأسماء الباحثين ومؤسساتهم من المستندات عبر الإنترنت وأعلنت عن مهمة تحديد عناوين بريدهم الإلكتروني على موقع إم تورك - نظام التعهيد الجماعي لشركة أمازون ميكانيكال تورك Amazon Mechanical Turk، فعناوين البريد الإلكتروني لمعظم الباحثين مدرجة على مواقع جامعاتهم. وبعد 24 ساعة و 54 دولاراً، كنت المالك الفخور لقائمة بريدية تضم مئات من الباحثين في الذكاء الاصطناعي الناجحين في أعمالهم بما يكفي ليُنتخبوا كزملاء لجمعية تقدم الذكاء الاصطناعي Association for the Advancement of Artificial Intelligence (اختصاراً: الجمعية AAAI)، أحدهم كان أستاذ الذكاء الاصطناعي البريطاني الأسترالي توبي والش Toby Walsh الذي وافق على إرسال بريد إلكتروني إلى أي شخص آخر في القائمة والمساعدة على قيادة حملتنا. وأنتج موظفو إم تورك حول العالم بلا كلل قوائم بريدية إضافية لتوبي، وقبل مضي وقت طويل، وقع أكثر من 3000 باحث في مجال الذكاء الاصطناعي والروبوتات على رسالتنا المفتوحة، بما في ذلك ستة من الرؤساء السابقين لجمعية AAAI وقادة صناعة الذكاء الاصطناعي من غوغل وفيسبوك وميكروسوفت وتسلا. وعمل جيش من متطوعي المعهد FLI بلا كلل من أجل التحقق من صحة قوائم الموقعين، ومسح الرسائل الهزلية التي تدعي أنها من بيل كلينتون Bill Clinton وسارة كونور Sarah Connor. ووقع أكثر من 17,000 شخص آخر أيضاً، بمن في ذلك ستيفن هوكينغ Stephen Hawking. وبعد أن نظم توبي مؤتمرا صحفيا حول هذا الموضوع في المؤتمر الدولي المشترك للذكاء الاصطناعي، صار الموضوع قصة إخبارية رئيسية في جميع أنحاء العالم.

نظراً لأن البيولوجيين والكيميائيين اتخذوا موقفاً من قبل، فمن المعروف الآن أن مجالاتهم العلمية معنية أساساً بإنتاج الأدوية والمواد المفيدة بدلاً من الأسلحة البيولوجية والكيميائية. كذلك أعربت مجتمعات الذكاء الاصطناعي والروبوتات أيضاً عن نفسها: أراد الموقعون على الخطابات أيضاً أن يُعرف مجالهم بخلق مستقبل أفضل، وليس لخلق طرق جديدة لقتل الأفراد، ولكن هل سيكون الاستخدام الرئيسي للذكاء الاصطناعي في المستقبل مدنيّاً أم عسكريّاً؟ على الرغم من أننا أنفقنا مزيداً من الصفحات في هذا الفصل على الجانب المدني، فقد ننفق قريباً المزيد من الأموال على الجانب العسكري - خاصةً إذا انطلق سباق التسلح العسكري للذكاء الاصطناعي. فقد تجاوزت الالتزامات الاستثمارية المدنية للذكاء الاصطناعي بليون دولار في عام 2016. ولكن هذا تضاءل أمام طلب ميزانية البنتاغون لعام 2017 الذي يتراوح بين 12 و15 بليون دولار للمشروعات ذات الصلة بالذكاء الاصطناعي، ومن المرجح أن تأخذ الصين وروسيا بعين الاعتبار ما قاله نائب وزير الدفاع روبرت وورك Robert Work عندما تم الإعلان عن ذلك: «أريد من منافسينا أن يتساءلوا عما وراء الستار الأسود».[41]

هل يجب أن تكون هناك معاهدة دولية؟

على الرغم من أن هناك الآن دفعة دولية كبيرة نحو التفاوض على شكل من أشكال حظر الروبوت القاتل، فإنه ما زال من غير الواضح ما الذي سيحدث. ويدور جدال مستمر حول ما يجب أن يحدث، هذا إذا كان يجب أن يحدث أي شي على الإطلاق، على الرغم من أن العديد من أصحاب المصلحة Stakeholders الرئيسيين يوافقون على أنه يجب على القوى العالمية صياغة بعض أشكال اللوائح الدولية لتوجيه أبحاث الأسلحة AWS واستخدامها. غير أن هناك اتفاقاً أقل حول ما يجب حظره على وجه الدقة وكيف سيتم فرض الحظر. مثلاً، هل يجب حظر الأسلحة الذاتية الفتاكة فقط، أم الأسلحة التي تؤذي الأفراد بشكل خطير، مثلاً بإصابتهم بالعمى؟ هل سنحظر التطوير أم الإنتاج أم الملكية؟ هل يجب أن ينطبق الحظر على جميع أنظمة الأسلحة الذاتية أو، كما ذكرنا، الأنظمة الهجومية فقط، ونسمح بأنظمة دفاعية مثل المدافع المضادة للطائرات الذاتية والدفاعات الصاروخية؟ في الحالة الأخيرة، هل يجب اعتبار الأسلحة AWS دفاعية حتى لو كان من السهل نقلها إلى أراضي العدو؟ وكيف يمكنك تطبيق معاهدة دولية مع أن معظم مكونات السلاح الذاتي لها استخدام مدني مزدوج أيضاً؟ مثلاً، لا يوجد فرق كبير بين طائرة من دون طيار يمكنها إيصال طرود الأمازون أو إيصال القنابل.

جادل بعض المتحاورين في أن تصميم معاهدة أسلحة AWS فعالة أمر صعب بشكل ميؤوس منه، وعلينا ألا نحاول ذلك. من ناحية أخرى، أكد جون إف. كينيدي John F. Kennedy عند الإعلان عن بعثات القمر Moon missions أن الأشياء الصعبة تستحق المحاولة عندما يكون النجاح مفيداً لمستقبل البشرية. إضافة إلى ذلك، يجادل العديد من الخبراء في أن الحظر المفروض على الأسلحة البيولوجية والكيميائية كان ذا قيمة، على الرغم من أن التطبيق صار صعباً نظراً للكثير من الغش، لأن حظرها وصمها بعار شديدة مما حدَّ من استخدامها.

وفي عام 2016 قابلت هنري كيسنجر Henry Kissinger في حفل عشاء، وأتيحت لي الفرصة لسؤاله عن دوره في حظر الأسلحة البيولوجية، وشرح كيف أنه عندما كان مستشار الأمن القومي الأمريكي تمكن من إقناع الرئيس نيكسون بأن الحظر سيكون مفيداً للأمن القومي الأمريكي. وقد تأثرت بمدى حدة ذهنه وذاكرته بالنسبة إلى شخص في التسعين من عمره، وقد فتنت بسماع وجهة نظره. ولما كانت الولايات المتحدة تتمتع بالفعل بوضع قوة عظمى بفضل قواتها التقليدية والنووية، فكانت ستخسر أكثر مما ستكسبه من سباق تسلح الأسلحة البيولوجية على مستوى العالم. بعبارة أخرى، إذا كنت الأعلى، فمن المنطقي اتباع الحد الأقصى، «إذا لم يكن مكسوراً فلا تُصلحه». انضم رسل ستيوارت إلى محادثتنا بعد العشاء، وناقشنا كيفية استخدام الحجة نفسها بالضبط حول الأسلحة الذاتية الفتاكة: أولئك الذين يستفيدون أكثر من سباق التسلح ليسوا قوى عظمى ولكنهم دول شريرة صغيرة والجهات الفاعلة جهات غير حكومية مثل الإرهابيين الذين يمكنهم الوصول إلى الأسلحة عبر السوق السوداء بمجرد تطويرها.

بمجرد إنتاجها بكميات كبيرة، من المرجح أن تكلف الطائرات الصغيرة القاتلة العاملة بالذكاء الاصطناعي أكثر قليلاً من سعر الهاتف الذكي. وسواء أكان ذلك إرهابيّاً يريد اغتيال سياسي أم عاشقاً مفجوعاً يسعى إلى الانتقام من صديقته السابقة، فكل ما يحتاجون إليه هو تحميل صورة وعنوان الهدف الخاص في الطائرة القاتلة: يمكن بعد ذلك الوصول إلى الوجهة وتحديد الهوية والقضاء عليها، ثم التدمير الذاتي لضمان ألّا أحد يعرف من كان المسؤول. أما بالنسبة إلى أولئك الذين يعتزمون التطهير العرقي، فيمكن برمجته بسهولة لقتل الأشخاص من لون أو عرق معين. ويتصور ستيوارت أنه كلما زادت هذه الأسلحة ذكاءً، قلت الحاجة إلى المواد والأموال اللازمة لكل عملية قتل. مثلاً، يخشى ستيوارت تطوير طائرات من دون طيار بحجم نحلة تقتل بثمن بخس باستخدام الحد الأدنى من القوة المتفجرة بإطلاق النار على الأفراد في العين التي هي لينة بما فيه الكفاية للسماح حتى لقذيفة صغيرة بأن تخترقها إلى الدماغ، أو قد تُطبق على الرأس بمخالب معدنية ثم تخترق الجمجمة بشحنة صغيرة الحجم. وإذا كان بالإمكان إطلاق مليون طائرة قاتلة من ظهر شاحنة واحدة، عندها تمتلك سلاحاً مرعباً من الدمار الشامل من نوع جديد تماماً: صاروخ يمكنه الاختيار انتقائيّاً من فئة محددة من الأفراد فقط، وترك البقية وكل شيء آخر سالماً.

وهناك حجة مضادة شائعة تتمثل في أنه يمكننا القضاء على مثل هذه المخاوف بجعل الروبوتات القاتلة أخلاقية - مثلاً، يقتلون جنود العدو فقط. ولكن إذا كُنّا قلقين بشأن فرض حظر، فكيف سيكون من الأسهل تطبيق شرط أن تكون أسلحة العدو الذاتية أخلاقية 100% بدلاً من أن يُفرض أنها لا تُنتج في المقام الأول؟ وقد يزعم المرء أن جنود الدول المتقدمة المدربين تدريباً جيداً سيئون جدّاً في اتباع قواعد الحرب وأن الروبوتات يمكنها القيام بذلك بشكل أفضل، في حين يدعي في الوقت نفسه أن الدول المارقة والديكتاتوريين والجماعات الإرهابية جيدة في اتباع قواعد الحرب؛ فلن يختاروا أبداً استخدام الروبوتات بطرق تنتهك هذه القواعد؟

حرب سيبرانية

الجانب العسكري الآخر المثير للاهتمام في الذكاء الاصطناعي هو أنه قد يسمح لك بمهاجمة عدوك حتى من دون بناء أي أسلحة خاصة بك، من خلال الحرب الإلكترونية. وكمقدمة صغيرة لما قد يجلبه المستقبل، فإن دودة *ستنكسنت* Stuxnet worm، التي تُنسب على نطاق واسع إلى الحكومتين الأمريكية والإسرائيلية، أصابت أجهزة الطرد المركزي في برنامج التخصيب النووي الإيراني وتسببت في تحطيمها لنفسها. كلما زادت أتمتة المجتمع زادت قوة الذكاء الاصطناعي، وزادت قدرة الحرب الإلكترونية على التدمير. فإذا تمكنت من اختراق وتحطيم سيارات العدو الذاتية القيادة، والطائرات ذات القيادة الآلية، والمفاعلات النووية، والروبوتات الاصطناعية، وأنظمة الاتصالات، والأنظمة المالية وشبكات الطاقة، فعندئذ يمكنك تحطيم اقتصاده وإعاقة دفاعاته، إذا كنت تستطيع اختراق بعض نظم أسلحته كذلك، فذلك أفضل وأفضل.

بدأنا هذا الفصل بدراسة المدى المذهل للفرص المتاحة على المدى القريب لفائدة الذكاء الاصطناعي للبشرية - إذا تمكنا من تطويره بمتانة وجعله غير قابل للاختراق. وعلى الرغم من أنه يمكن استخدام الذكاء الاصطناعي نفسه لجعل أنظمة الذكاء الاصطناعي أكثر متانة، ومن ثم يساعد على الدفاع ضد الحرب السيبرانية، ومن الواضح أن الذكاء الاصطناعي قادر على المساعدة على الهجوم أيضاً. ويجب ضمان أن يكون الدفاع هو الهدف الأهم من الأهداف قصيرة الأجل لتطوير الذكاء الاصطناعي - وإلا ستُقلب جميع التكنولوجيات الرائعة التي طورناها ضدنا!

وظائف وأجور

حتى الآن في هذا الفصل، ركزنا بشكل أساسي على كيفية تأثير الذكاء الاصطناعي فينا كمستهلكين، من خلال تمكين منتجات وخدمات تحويلية جديدة بأسعار معقولة، ولكن كيف ستؤثر فينا **كعاملين**، من خلال تحويل سوق العمل؟ إذا تمكنا من معرفة كيفية تنمية ازدهارنا من خلال الأتمتة دون حرمان الأفراد من الدخل أو الغرض، عندئذ تكون لدينا القدرة على خلق مستقبل رائع مع أوقات استجمام وثراء غير مسبوقين لكل مَنْ يريد ذلك. قليل من فكر في الأمر لفترة أطول وبشكل أعمق من الاقتصادي إريك برينجولفسون Erik Brynjolfsson، أحد زملائي في المعهد MIT. وعلى الرغم من أنه دائماً مرتب ويرتدي ملابس أنيقة، فهو يتمتع بتراث آيسلندي، فإنني أحياناً لا أستطيع إلاّ أن أتخيل أنه مؤخراً خفف لحية الفايكنغ الحمراء ولبدة شعره الجامحين ليتماشى مع أجواء كلية إدارة الأعمال. ومن المؤكد أنه لم يخُفف من أفكاره الجامحة، ويُطلق على رؤيته المتفائلة لسوق العمل أثينا الرقمية Digital Athens، والسبب في أن المواطنين الأثينيين في العصور القديمة كانوا يعيشون حياة استجمام مستمتعين بالديمقراطية والفن والألعاب لأنه كان لديهم عبيد للقيام بالكثير من العمل. لكن لماذا لا تحل محل العبيد روبوتات قائمة على الذكاء الاصطناعي، ونخلق يوتوبيا رقمية Digital utopia يمكن للجميع الاستمتاع بها؟ إن اقتصاد إيريك الذي يحركه الذكاء الاصطناعي لن يلغي فقط الإجهاد والكدح ويؤدي إلى وفرة من كل ما نريده اليوم، ولكنه سيوفر أيضاً مجموعة كبيرة من المنتجات والخدمات الجديدة الرائعة التي لم يدركها مستهلكو اليوم حتى الآن.

التكنولوجيا وعدم المساواة

يمكننا أن ننتقل من حيث نحن اليوم إلى أثينا إيريك الرقمية Erik's Digital Athens إذا استمر مُرتَّب كل ساعة في النمو عاماً بعد عام، حتى يتسنى تدريجيّاً لأولئك الذين يريدون المزيد من الاستجمام العمل لفترات أقل مع الاستمرار في تحسين مستوى معيشتهم. ويوضح الشكل 5.3 أن هذا هو بالضبط ما حدث في الولايات المتحدة منذ الحرب العالمية الثانية حتى منتصف سبعينات القرن العشرين: على الرغم من عدم مساواة في الدخل، فإن

الحجم الإجمالي للفطيرة نما بمثل هذه المنوال وحصل الجميع على شريحة أكبر. لكن بعد ذلك، وإريك هو أول المعترفين بذلك، تغير شيء ما: الشكل 5.3 يوضح أنه على الرغم من أن الاقتصاد متوسط الدخل استمر بالنمو، فإن المكاسب التي تحققت على مدار العقود الأربعة الماضية كانت من نصيب الأكثر ثراءً من الشريحة الأعلى ثراء والتي تمثل 1%، في حين أن الأفقر -90%- رأوا دخولهم تراوح محلها. والنمو الناتج في عدم المساواة يصبح أكثر وضوحاً إذا لم ننظر إلى الدخل بل إلى الثروة، بالنسبة إلى 90% الأدنى من الأسر في الولايات المتحدة، كان متوسط القيمة الصافية نحو 85,000 دولار في عام 2012 - مثلما كان المتوسط قبل 20 عاماً، في حين ضاعفت الـ 1% الأعلى ثرواتها -المعدلة حسب التضخم- خلال تلك الفترة إلى 14 مليون دولار.[42] بل إن الاختلافات أكثر تطرفاً على المستوى الدولي، ففي عام 2013 كانت ثروة نصف الشريحة الأدنى من سكان العالم مجتمعة (أكثر من 3.6 بليون شخص) تعادل ثروات أغنى ثمانية أشخاص في العالم [43] - بيانات إحصائية تبرز مدى الفقر والتعرض للمخاطر في القاع ومقدار الثروة في القمة. وفي مؤتمر بورتوريكو لعام 2015، أخبر إريك باحثي الذكاء الاصطناعي المجتمعين بأنه يعتقد أن التقدم في الذكاء الاصطناعي والتشغيل الآلي سيستمران في زيادة حجم الكعكة الاقتصادية. لكن لا يوجد قانون اقتصادي يُحتم أن يستفيد الجميع، أو حتى معظم الأشخاص.

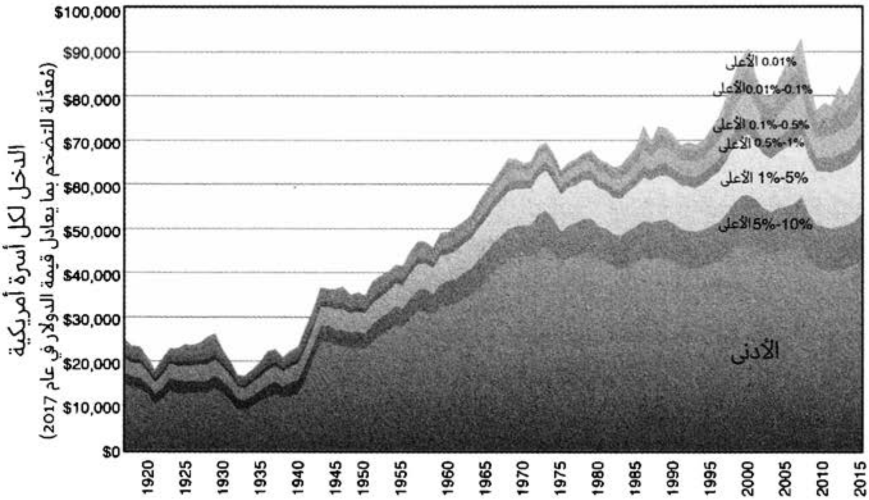
وعلى الرغم من وجود اتفاق واسع بين الاقتصاديين على أن عدم المساواة آخذ بالازدياد، فإن هناك جدلاً مثيراً للاهتمام حول سبب هذا الاتجاه وما إذا كان سيستمر. غالباً ما يقول المجادلون ممن هم على الجانب الأيسر من المدى السياسي إن السبب الرئيسي هو العولمة و/أو السياسات الاقتصادية مثل التخفيضات الضريبية للأثرياء. لكن إريك برينجولفسون وزميله من المعهد MIT أندرو مكافي Andrew McAfee يجادلان في أن السبب الرئيسي هو شيء آخر: التكنولوجيا.[44] على وجه التحديد، يجادلان في أن التكنولوجيا الرقمية تدفع بعدم المساواة بثلاث طرق مختلفة.

أولاً. من خلال الاستعاضة عن الوظائف القديمة بوظائف تتطلب المزيد من المهارات. فقد كافأت التكنولوجيا المتعلمين: منذ منتصف سبعينات القرن العشرين، ارتفعت رواتب الحاصلين على شهادات عليا بنحو 25%، في حين انخفضت أجور المتسربين من المدارس الثانوية في المتوسط بنسبة 30% في الأجور.[45]

ثانياً. يزعمان أنه منذ عام 2000 ذهبت حصة أكبر من دخل الشركات لأولئك الذين يمتلكون الشركات بدلاً من أولئك الذين يعملون فيها - وما دامت الأتمتة مستمرة، فيجب أن نتوقع أن من يمتلك الآلات سيأخذ جزءاً متزايداً من الكعكة. قد تكون هذه الميزة لرأس المال على العمالة مهمة بشكل خاص في الاقتصاد الرقمي المتنامي، الذي يُعرّفه نيكولاس نغروبونتي Nicholas Negroponte ذو الرؤية التكنولوجية بمثابة البتات المتحركة، وليس الذرات.

الآن وقد صار كل شيء -بدءاً من الكتب وحتى الأفلام وأدوات إعداد الإقرار الضريبي- رقميّاً، يمكن بيع نسخ إضافية في جميع أنحاء العالم بتكلفة تعادل الصفر تقريباً، دون تعيين موظفين إضافيين. ويتيح ذلك أن تذهب معظم العوائد إلى المستثمرين بدلاً من العمال،

ويساعد ذلك على توضيح السبب في سبب كون الإيرادات المجمعة لشركات ديترويت الثلاث الكبار "Big 3" (جنرال إلكتريك GM وفورد Ford وكرايسلر Chrysler) في عام 1990 مماثلة تقريباً لإيرادات الثلاثة الكبار في سيليكون فالي Silicon Valley (غوغل Google وآبل Apple وفيسبوك Facebook) في عام 2014، وكان عدد الموظفين في شركات سيليكون فالي أقل بتسعة أضعاف وكان سعرها أعلى بنحو 30 مرة في سوق الأوراق المالية.[47]



الشكل 5.3: كيف دفع الاقتصاد بمتوسط الدخل على مدى القرن الماضي، وأي جزء من هذا الدخل ذهب إلى المجموعات المختلفة. قبل سبعينات القرن العشرين كانت دخول الأغنياء والفقراء جميعاً تتحسن بتناسب. بعد ذلك استحوذت الشريحة العليا التي تمثل 1% من السكان على معظم المكاسب في حين لم تكسب الشريحة الدنيا التي تمثل 90% أي شيء تقريباً.[46] -الأعداد مُعادلة إلى قيمة الدولار في عام 2017.

ثالثاً، يجادل إريك والمتعاونون معه في أن الاقتصاد الرقمي يُفيد النجوم دون سواهم، فقد صارت مؤلفة كتاب هاري بوتر Harry Potter المؤلفة جي. كي. رولينغ J. K. Rowling أول كاتبة تنضم إلى نادي البليونيرات، وصارت أكثر ثراءً من شكسبير لأن قصصها يُمكن أن تُحوَّل من نص إلى أفلام وألعاب لبلايين الأشخاص بتكلفة منخفضة جدّاً. وبالمثل، جنى سكوت كوك Scott Cook بليون دولار من تيربو تاكس TurboTax - البرنامج الحاسوبي لإعداد الإقرار الضريبي. فعلى عكس معدي الضرائب من البشر، يمكن بيعه كتحميل من الإنترنت، ولما كان معظم الأفراد على استعداد لدفع القليل أو لا شيء مقابل البرنامج الحاسوبي العاشر من العشرة الأوائل من برمجيات إعداد الإقرار الضريبي، فهناك مساحة لعدد متواضع فقط من النجوم في السوق، هذا يعني أنه إذا نصح جميع آباء العالم أطفالهم بأن يصبحوا جي. كي. رولينغ، أو جيزيل بوندشين Gisele Bündchen، أو مات دامون Matt Damon، أو كريستيانو رونالدو Cristiano Ronaldo، أو أوبرا وينفري nfrey أو إيلون ماسك Elon Musk، التالين فبالكاد سيجد أطفالهم أن هذه استراتيجية مهنية قابلة للتطبيق.

المشورة المهنية للأطفال

إذاً، ما النصيحة المهنية التي يجب أن نقدمها إلى أطفالنا؟ أنا أشجعُ أطفالي على الخوض في المهن التي تكون الآلات فيها سيئة حاليّاً، ويبدو أنه من غير المرجح أن تؤتمت في المستقبل القريب. وتحدد التنبؤاتُ الحديثة بموعد استيلاء الآلات على الوظائف المختلفة العديد من الأسئلة المفيدة التي يجب طرحها حول مهنة ما قبل اتخاذ قرار تتبع ذلك المسار،[48] مثلاً:

- هل تتطلب هذه المهنة التفاعلَ مع الناس واستخدام الذكاء الاجتماعي؟
- هل تنطوي هذه المهنة على الإبداع والتوصل إلى حلول ذكية؟
- هل تتطلب هذه المهنة العمل في بيئة غير متوقعة؟

كلما زادت الأسئلة التي يمكنك الإجابة عنها بنعم، كان اختيارك الوظيفي أفضل، هذا يعني أن الرهانات الآمنة نسبيّاً تشمل أن تصبح مدرساً أو ممرضة أو طبيباً أو طبيب أسنان أو عالماً أو رائد أعمال أو مبرمجاً أو مهندساً أو محامياً أو اختصاصياً اجتماعياً أو رجل دين أو فناناً أو مصفف شعر أو معالج تدليك.

وفي المقابل، فإن الوظائف التي تنطوي على إجراءات متكررة جدّاً أو منظمة في بيئة يمكن التنبؤ بها، من غير المحتمل أن تستمر لفترة طويلة قبل أن تؤتمت. فقد استحوذت أجهزة الحاسوب والروبوتات الصناعية على أبسط هذه الوظائف منذ فترة طويلة. كما أن تحسين التكنولوجيا في طريقه إلى القضاء على العديد من الوظائف الأخرى، من التسويق عبر الهاتف إلى عمال المستودعات والصرافين ومشغلي القطارات والخبازين و طهاة المطاعم.[49] ومن المحتمل أن يتبعهم قريبا سائقو الشاحنات والحافلات وسيارات الأجرة وسيارات ليفت/أوبر. وهناك العديد من المهن الأخرى (بمن في ذلك المساعدون القانونيون ومحللو الائتمان وموظفو القروض والمحاسبون ومحاسبو الضرائب) الذين على الرغم من أنهم غير مدرجين في القائمة المهددة بالانقراض التام، فإن معظم مهامهم تلقائية ومن ثم تتطلب عدداً أقل من الأفراد.

وعلى الرغم من أن البقاء بعيداً عن الأتمتة ليس هو التحدي الوظيفي الوحيد، في هذا العصر الرقمي العالمي، فإن السعي إلى أن تصبح كاتباً أو صانع أفلام أو ممثلاً أو رياضياً أو مصمماً للأزياء أمر محفوف بالمخاطر لسبب آخر: على الرغم من أن الأشخاص في هذه المهن لن يجابهوا منافسة جادة من الآلات في أي وقت قريب، فإنهم سيصبحون أكثر فأكثر معرضين للمنافسة الشرسة من البشر الآخرين في جميع أنحاء العالم وفقاً لنظرية النجم السالفة الذكر، ولن ينجح منهم إلا القليل.

وفي كثير من الحالات سيكون من قصر النظر والسذاجة إسداء نصيحة مهنية على مستوى جميع المجالات: فهناك العديد من الوظائف التي لن تنقرض بالكامل، ولكن سنرى العديد من مهامها مؤتمتة. مثلاً، إذا دخلت مجال الطب، فلن تكون اختصاصي

الأشعة الذي يحلل الصور الطبية ويحل محله واتسون من شركة IBM، ولكن الطبيب الذي يطلب تحليل الأشعة، ويناقش النتائج مع المريض، ويقرر خطة العلاج. إذا دخلت مجال التمويل، فلن تكون "الضليع في الرياضيات" الذي يطبق الخوارزميات على البيانات فيُستبدل بالبرمجيات، ولكن مدير المحفظة Fund manager الذي يستخدم نتائج التحليل الأساسي لاتخاذ قرارات استثمارية استراتيجية. وإذا دخلت في القانون، فلن تكون المساعد القانوني الذي يراجع الآلاف من المستندات خلال مرحلة الاستكشاف ويؤتمت، ولكن الشخص الذي ينصح العميل ويترافع في القضية أمام المحكمة.

لقد اكتشفنا حتى الآن ما يمكن للأفراد فعله لزيادة نجاحهم في سوق العمل في عصر الذكاء الاصطناعي، ولكن ما الذي تستطيع الحكومات فعله لمساعدة القوى العاملة لديها على النجاح؟ مثلاً، ما هو النظام التعليمي الذي يعد الأشخاص على أفضل وجه لسوق عمل يستمر فيها تحسن الذكاء الاصطناعي تحسناً سريعاً؟ هل لا يزال هو نفسه نموذجنا الحالي مع عقد أو عقدين من التعليم يتبعه أربعة عقود من العمل المتخصص؟ أم أنه من الأفضل التبديل إلى نظام يعمل فيه الأشخاص لبضع سنوات، ثم يعودون إلى المدرسة لمدة عام، ثم العمل لبضع سنوات أخرى؟[50] أم هل ينبغي أن يكون التعليم المستمر (ربما يتم توفيره عبر الإنترنت) جزءاً قياسياً من أي عمل؟

وما هي السياسات الاقتصادية الأكثر فائدة لخلق وظائف جديدة جيدة؟ يجادل أندرو مكافي في أن هناك العديد من السياسات التي قد تساعد، بما في ذلك الاستثمار بكثافة في البحث والتعليم والبنية التحتية، وتسهيل الهجرة وتحفيز ريادة الأعمال، إنه يشعر بأن «كتاب مقرر اقتصاد 101 Econ 101 هو منهج واضح. لكنه لا يُتبع»، على الأقل ليس في الولايات المتحدة.[51]

هل سيغدو البشر في نهاية المطاف عاطلين عن العمل؟

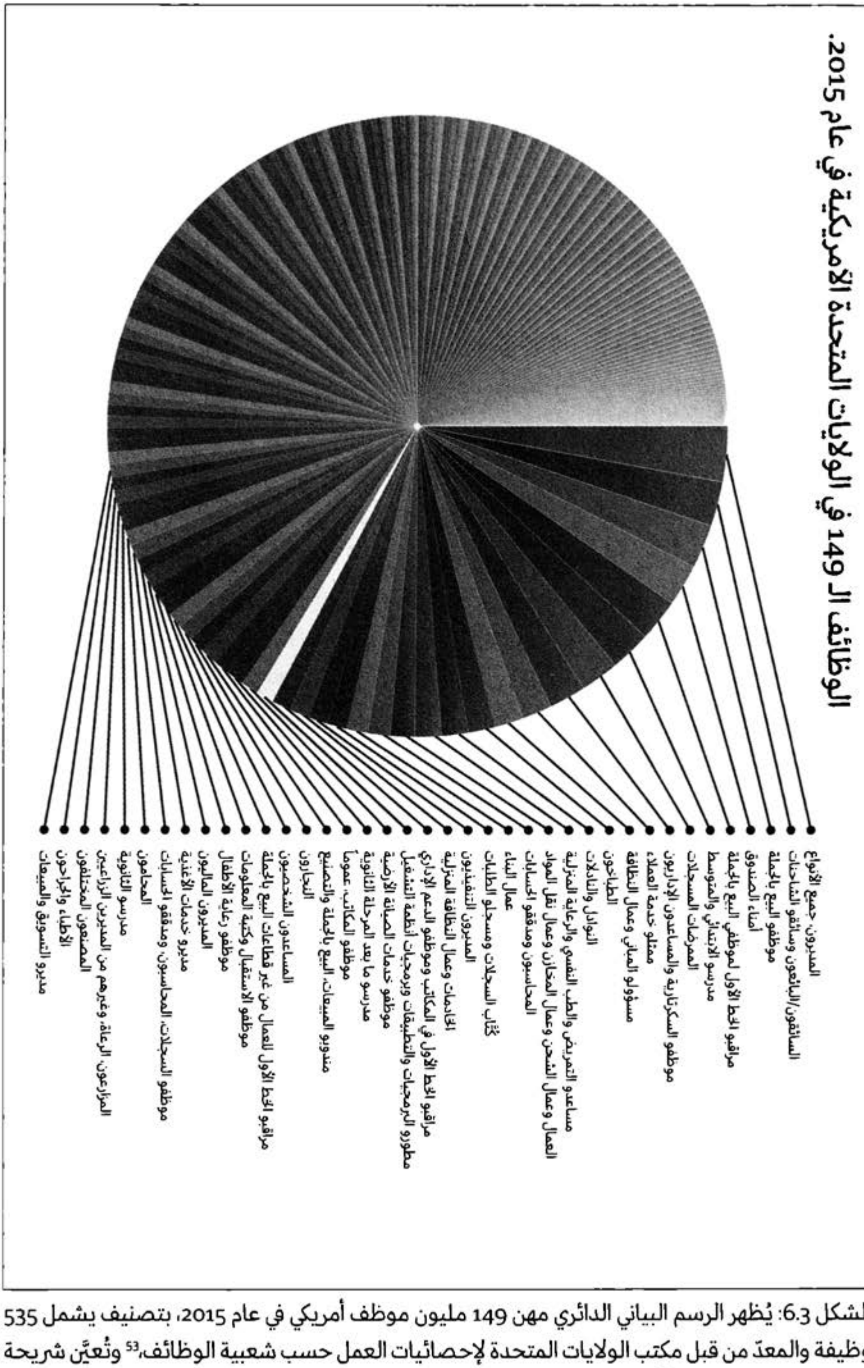
إذا استمر الذكاء الاصطناعي في التحسن، وأتمتة المزيد من الوظائف، فما الذي سيحدث؟ كثير من الناس متفائلون من حيث الوظائف، بحجة أن الوظائف المؤتمتة ستُستبدل بأخرى جديدة أفضل، فهذا ما كان يحدث دائماً من قبل، منذ أن كان اللاضيون (أعاد الآلة) Luddites يساورهم القلق بشأن البطالة التقنية أثناء الثورة الصناعية.

ولكن، هناك آخرون متشائمون من حيث الوظيفة ويجادلون في أن الأمر هذه المرة مختلف تماماً، وأن عدداً أكبر من الأفراد لن يصبحوا العاطلين عن العمل فقط ولكن غير قادرين على العمل.[52] ويجادل المتشائمون من حيث مستقبل الوظائف في أن السوق الحرة تحدد الرواتب بناءً على العرض والطلب، وأن العرض المتزايد للعمالة الرخيصة سيؤدي في النهاية إلى خفض رواتب الأفراد إلى ما هو أقل كثيراً عن تكلفة المعيشة، نظراً لأن رواتب السوق بالنسبة إلى الوظيفة هي التكلفة بالساعة لمن أو أياً كان من شأنه أن يؤديها

بأرخص الأسعار. فقد انخفضت الرواتب تاريخيّاً كلما صار من الممكن الاستعانة بمصادر خارجية لمهنة معينة من بلد منخفض الدخل أو إلى آلة رخيصة. وأثناء الثورة الصناعية بدأنا باستكشاف كيفية الاستعاضة عن عضلاتنا بالآلات، وانتقل الأفراد إلى وظائف ذات رواتب أفضل؛ إذ استخدموا عقولهم أكثر، واستعيض عن وظائف ذوي الياقات الزرقاء (العمال اليدويين) Blue-collar بوظائف ذوي الياقات البيضاء (الموظفين المكتبيين) White-collar، نحن الآن ندرس تدريجيّاً كيفية الاستعاضة عن عقولنا بالآلات، إذا نجحنا في النهاية في هذا، فما الوظائف التي ستتبقى لنا؟

يجادل بعض المتفائلين بمستقبل الوظائف في أنه بعد الوظائف الجسدية والعقلية، ستكون الطفرة التالية في الوظائف الإبداعية. ولكن المتشائمين بمستقبل الوظائف يعارضون أن الإبداع هو مجرد عملية عقلية أخرى، سيتقنها الذكاء الاصطناعي في النهاية الأمر أيضاً. ويأمل متفائلون آخرون بمستقبل الوظائف أن تكون الطفرة التالية في مهن جديدة معتمدة على التكنولوجيا لم نفكر فيها حتى الآن. فمن كان يتخيل خلال الثورة الصناعية أن أحفادهم سيعملون يوماً ما كمصممين للويب وسائقي أوبر؟ ولكن المتشائمين بمستقبل الوظائف يرفضون هذا التفكير لأنه عبارة عن أُمنيات؛ مما قد يدعمها القليل جدّاً من البيانات التجريبية، ويشيرون إلى أنه كان بإمكاننا أن نُجادل في استخدام الحجة نفسها قبل قرن من الزمن، قبل ثورة الحاسوب، وكنا سنتوقع أن معظم مهن اليوم ستكون جديدة ولم تكن متخيلة في السابق وتدعم تكنولوجيا لم تكن موجودة. وكان هذا التوقع بمثابة فشل ذريع، كما هو موضح في الشكل 6.3: الغالبية العظمى من المهن الحالية هي تلك التي كانت موجودة بالفعل منذ قرن مضى. وعندما نفرزها حسب عدد الوظائف التي توفرها، علينا أن ننزل جميعاً إلى المرتبة الحادية والعشرين في المرحلة الأولى حتى نصادف وظيفة جديدة: مطورو البرمجيات الذين يشكلون أقل من 1% من سوق العمل في الولايات المتحدة.

يمكننا الحصول على فهم أفضل لما يحدث باستحضار الشكل 2.2 من الفصل الثاني. والذي أظهر مشهد الذكاء البشري، مع ارتفاع يمثل كيف أنه من الصعب للآلات أداء مختلف المهام، ومدى ارتفاع مستوى سطح البحر يظهر ما يمكن آلات القيام به حاليّاً. والاتجاه الرئيسي في سوق العمل هو أننا لا ننتقل إلى مهن جديدة تماماً، بدلاً من ذلك، نحن نحتشد في تلك الأجزاء من التضاريس في الشكل 2.2 الذي لم تغمره الموجة التكنولوجية المتصاعدة بعد! يوضح الشكل 6.3 أن هذا لا يشكل جزيرة واحدة بل أرخبيلاً معقداً، مع جزر وأتولات مرجانية Atolls تُمثِّل جميع الأشياء القيمة التي لا تزال الآلات غير قادرة على القيام بها بثمن أرخص مما يستطيعه البشر، ولا يشمل هذا المهن التقنية العالية فقط مثل تطوير البرمجيات، ولكن يشمل أيضاً مجموعة كبيرة من الوظائف ذات التقنية المنخفضة التي تستفيد من مهارتنا الفائقة ومهاراتنا الاجتماعية، بدءاً من العلاج بالتدليك وحتى التمثيل. هل سيتفوَّق علينا الذكاء الاصطناعي في المهام الفكرية بسرعة كبيرة، بحيث تكون آخر الوظائف المتبقية هي هذه الفئة ذات التكنولوجيا المنخفضة؟



الشكل 6.3: يُظهر الرسم البياني الدائري مهن 149 مليون موظف أمريكي في عام 2015، بتصنيف يشمل 535 وظيفة والمعدّ من قبل مكتب الولايات المتحدة لإحصائيات العمل حسب شعبية الوظائف،[53] وتُعيَّن شريحة تصنيفية لكل مهنة تضم أكثر من مليون عامل، ولا توجد مهن جديدة تطورت بفعل تكنولوجيا الحاسوب حتى نصل إلى المرتبة الحادي والعشرين. هذه الفكرة مستوحاة من تحليل فديريكو بيستونو Federico Pistono.[54]

يخلص المتشائمون بمستقبل الوظائف إلى أن نقطة النهاية واضحة: سيُغمر الأرخبيل بكامله، ولن تبقى هناك وظائف يمكن للبشر أداؤها بثمن أدنى من الآلات. وفي كتابه الصادر عام 2007 بعنوان **وداعاً للصدقات *Farewell to Alms***، يشير الاقتصادي الأمريكي الأسكتلندي غريغوري كلارك Gregory Clark إلى أن بإمكاننا تعلم شيء أو اثنين عن فرص عملنا المستقبلية من خلال مقارنة ذلك بأصدقائنا من الخيول، تخيل اثنين من الخيول ينظران إلى سيارة مبكرة في عام 1900 ويفكران في مستقبلهما.

«أنا قلق بشأن البطالة التكنولوجية».

«نيه، نيه، لا تكن لاضيا (رافضا للآلات)، فقد قال أسلافنا القول نفسه عندما استولت المحركات البخارية على وظائف صناعتنا واستولت القطارات على وظائفنا في جر عربات السفر. ولكن لدينا اليوم وظائف أكثر من أي وقت مضى، وهي أفضل أيضاً: أفضل سحب عربة خفيفة عبر المدينة بدلاً من قضاء يوم كامل في المشي في دوائر لتشغيل مضخة بئر المنجم».

«لكن ماذا لو انتشر حقّاً هذا المحرك الذي يعمل بالاحتراق الداخلي؟»

«أنا متأكد من أنه ستكون هناك وظائف جديدة للخيول لم نتخيلها بعد، هذا ما حدث دائماً من قبل، كما هي الحال مع اختراع العجلة والمحراث».

للأسف، لم تظهر تلك الوظائف الجديدة غير المتخيلة بعد للخيول، وذبحت الخيول التي لم تعد لنا حاجة إليها ولم تُستبدل بأخرى جديدة؛ مما تسبب في انهيار عدد الخيول الأمريكية من نحو 26 مليوناً في عام 1915 إلى نحو ثلاثة ملايين في عام 1960[55]. ومثلما جعلت العضلاتُ الاصطناعية الخيولَ زائدة عن الحاجة، فهل ستؤدي العقول الميكانيكية إلى نتيجة مشابهة في البشر؟

منح الأفراد دخلا دون وظائف

إذن، من هو الذي على حق: أولئك الذين يقولون إن الوظائف المؤتمتة سيستعاض عنها بأخرى أفضل منها، أم أولئك الذين يقولون إن معظم البشر سينتهي بهم المطاف إلى أن يكونوا عاطلين عن العمل؟ إذا استمر تقدم الذكاء الاصطناعي بلا هوادة، فقد يكون كلا الطرفين على حق: أحدهما على المدى القصير والآخر على المدى الطويل. ولكن على الرغم من أن الأشخاص غالباً ما يناقشون اختفاء الوظائف بدلاً من الهلاك والكآبة، فإن الأمر لا يتعين أن يكون شيئاً سيئاً! مشغولون بهاجس وظائف معينة أهمل اللاضيون (رافضوا الآلات) احتمال أن وظائف أخرى قد توفر القيمة الاجتماعية نفسها. بالمثل، ربما يكون أولئك المشغولون بهاجس الوظائف اليوم ضيقي الأفق جدّاً: فنحن نريد وظائف لأنها توفر لنا الدخل والغرض، ولكن بالنظر إلى وفرة الموارد التي تنتجها الآلات، ينبغي أن يكون من الممكن إيجاد طرق بديلة لتوفير كل من الدخل والغرض من دون وظائف. انتهى الأمر

بشيء مماثل في قصة الخيول، التي لم تنته بانقراض جميع الخيول، بدلاً من ذلك، تضاعف عدد الخيول أكثر من ثلاثة أضعاف منذ عام 1960، إذ كانت محمية من قبل نوع ما من نظام رعاية اجتماعية للخيول: على الرغم من أنها لا تستطيع دفع الفواتير الخاصة بها، قرر الناس رعاية الخيول، والحفاظ عليها للمرح والرياضة والرفقة، هل يمكننا بالمثل رعاية زملائنا البشر المحتاجين؟

لنبدأ بمسألة الدخل: إعادة توزيع مجرد جزء صغير من الكعكة الاقتصادية المتنامية يجب أن يمكّن الجميع من أن يصبحوا في وضع أفضل، يجادل كثيرون في أننا لا نستطيع فحسب بل يجب أن نفعل هذا، في لجنة في عام 2016، عندما تحدث موشيه فاردي Moshe Vardi عن ضرورة أخلاقية لإنقاذ الأرواح بالتكنولوجيا المدعومة بالذكاء الاصطناعي، جادلتُ -أنا- في أنه من الضروري أخلاقيّاً الدعوة إلى استخداماتها المفيدة، بما في ذلك تقاسم الثروة، وقال إريك برينجولفسون، وهو أيضاً عضو في اللجنة: «مع كل هذا الإنتاج الجديد من الثروة، عارٌ علينا إذا لم نتمكن حتى من منع نصف جميع الناس من أن يصبحوا أسوأ حالاً!»

هناك العديد من المقترحات المختلفة لتقاسم الثروة، لكل منها مؤيدوها ومنتقدوها، أبسطها هو الدخل الأساسي. إذ يتلقى كل شخص دفعات شهرية دون أي شروط مسبقة أو متطلبات على الإطلاق، يُجرى الآن عدد من التجارب الصغيرة أو التخطيط لها، مثلاً في كندا وفنلندا وهولندا، ويزعم المدافعون أن الدخل الأساسي أكثر فعالية من البدائل مثل معونات الرعاية الاجتماعية للمحتاجين، لأنه يلغي المتاعب الإدارية المتمثلة في تحديد من هم المؤهلون، كما انتُقدت معونات الرعاية الاجتماعية الحالية لأنها لا تُحفّز على العمل، ولكن هذا بالطبع يصبح غير ذي صلة في مستقبل العاطلين عن العمل، إذ لا يعمل أحد.

هذا، وتستطيع الحكومات مساعدة مواطنيها ليس فقط من خلال منحهم المال، ولكن أيضاً من خلال تزويدهم بخدمات مجانية أو مدعومة مثل الطرق والجسور والحدائق العامة والنقل العام ورعاية الأطفال والتعليم والرعاية الصحية ودور التقاعد والوصول إلى الإنترنت؛ في الواقع، العديد من الحكومات توفر بالفعل معظم هذه الخدمات. وعلى العكس من الدخل الأساسي، تحقق هذه الخدمات التي تمولها الحكومة هدفين منفصلين: تخفيض تكلفة معيشة الأفراد، وتوفير الوظائف. وحتى في المستقبل الذي يمكن أن تتفوق فيه الآلات على البشر في جميع الوظائف، يمكن للحكومات أن تختار دفع أجر للأفراد للعمل في رعاية الأطفال، ورعاية المسنين، وما إلى ذلك بدلاً من تعهد تقديم الرعاية للروبوتات.

ومن المثير للاهتمام، أن التقدم التكنولوجي قد ينتهي إلى توفير العديد من المنتجات والخدمات القيمة مجاناً حتى من دون تدخل الحكومة. مثلاً، اعتاد الأشخاص على الدفع مقابل الموسوعات والأطالس وإرسال الرسائل وإجراء المكالمات الهاتفية، ولكن الآن أي شخص لديه اتصال بالإنترنت يمكنه الوصول إلى كل هذه الأشياء دون أي تكلفة - إضافة

إلى مؤتمرات الفيديو المجانية وتقاسم الصور والوسائط الاجتماعية والدورات التدريبية عبر الإنترنت و عدد لا يحصى من الخدمات الجديدة الأخرى. والعديد من الأشياء الأخرى التي قد تكون ذات قيمة عالية لأي شخص، مثلاً علاج مضادات حيوية منقذ للحياة، صار رخيصاً جدّاً. لذلك بفضل التكنولوجيا، يمكن حتى للعديد من الفقراء اليوم الوصول إلى الأشياء التي كان يفتقر إليها أغنى أغنياء العالم في الماضي، ويرى البعض أن هذا يعني أن الدخل اللازم لحياة كريمة آخذ بالانخفاض.

إذا تمكنت الآلات في يوم من الأيام من إنتاج جميع السلع والخدمات الحالية بأقل تكلفة، فمن الواضح أن هناك ثروة كافية لجعل الجميع في وضع أفضل. بعبارة أخرى، حتى الضرائب المتواضعة نسبيّاً يمكن أن تسمح للحكومات بدفع دخل أساسي وتوفير خدمات مجانية. ولكن حقيقة أن تقاسم الثروة يمكن أن يحدث لا تعني أنه سيحدث. وهناك اليوم خلاف سياسي قوي حول ما إذا كان يجب أن يحدث. كما رأينا فيما سبق، يبدو أن الوضع الحالي في الولايات المتحدة يمضي في الاتجاه المعاكس، إذ صارت بعض الفئات أكثر فقراً مما كانت قبل عقد من الزمن. وستؤثر القرارات السياسية حول كيفية تقسيم ثروة المجتمع المتنامية في الجميع. ومن ثم، فإن الحوار حول أي نوع من الاقتصاد المستقبلي يجب أن يشمل الجميع، وليس مجرد الباحثين في الذكاء الاصطناعي، ومطوري الروبوتات، والاقتصاديين.

يجادل الكثيرون في أن الحد من عدم المساواة في الدخل فكرة جيدة ليست فقط في مستقبل يهيمن عليه الذكاء الاصطناعي، ولكن اليوم أيضاً. وعلى الرغم من أن الحجة الرئيسية تميل إلى أن تكون حجة أخلاقية، فهناك أيضاً أدلة على أن تحقيق قدر أكبر من المساواة يجعل الديمقراطية تعمل بشكل أفضل: عندما تكون هناك طبقة وسطى كبيرة متعلمة تعليماً جيداً، يصعب على الناخبين التلاعب بها، ويصبح من الصعب على عدد صغير من الأفراد أو الشركات الحصول على قدرة تأثير -غير مستحقة- في الحكومة. ويمكن لديمقراطية أفضل بدورها أن تمكّن اقتصاداً مُداراً إدارة أفضل؛ فيكون أقل فساداً وأكثر فاعلية وأسرع نمواً، وفي النهاية يستفيد منه الجميع.

إعطاء الأفراد غرضاً من دون وظائف

الوظائف قد توفر للأفراد ما هو أكثر من مجرد أموال، وكتب فولتير في عام 1759: «إن العمل يحول دون ثلاثة شرور عظيمة: السأم، والرذيلة، والحاجة». وعلى العكس من ذلك، فإن توفير دخل للأشخاص لا يكفي لضمان رفاهيتهم. فقد وفّر الأباطرة الرومان كلاً من الخبز والسيرك للحفاظ على رضا رعيتهم. وشدد «يسوع» على الاحتياجات غير المادية في اقتباس من الكتاب المقدس «ليس بالخبز وحده يحيا الإنسان». إذاً، بالضبط ما الأشياء القيمة التي توفرها الوظائف فيما عدا المال، وبأي طرق بديلة يمكن لمجتمع العاطلين عن العمل أن يوفرها لهم؟

من الواضح أن الإجابات عن هذه الأسئلة معقدة، لأن البعض يكرهون وظائفهم ويحبها آخرون. إضافة إلى ذلك، فإن العديد من الأطفال والطلبة وربات البيوت يزدهرون من دون وظائف، في حين يعجّ التاريخ بقصص من الورثة المدللين والأمراء الذين استسلموا للسأم والاكتئاب. أظهر تحليل تلوي Meta-analysis أُجري في عام 2012 أن البطالة تميل إلى أن تكون لها آثار سلبية طويلة الأجل في الرفاهية، في حين أن التقاعد كان وعاء مختلطاً من الجوانب الإيجابية والسلبية.[56] وحدد مجال **علم النفس الإيجابي *Positive psychology*** المتنامي عدداً من الوجوه الخلل الذي يعزز إحساس الأفراد بالرفاهية والغرض، ووجد أن بعض الوظائف (وليس كلها!) قد توفر الكثير، مثلاً:[57]

- شبكة اجتماعية من الأصدقاء والزملاء
- نمط حياة صحي وفاضل
- الاحترام، وتقدير الذات، والشعور بالكفاءة، وشعور ممتع بـ"التدفق" Flow النابع من القيام بشيء يجيده الفرد
- الشعور بأن هناك حاجة إلى الفرد والقدرة على إحداث فرق
- الشعور بالمعنى النابع من كونه جزءاً من شيء أكبر من نفسه

هذا يبعث على التفاؤل، إذ يمكن توفير كل هذه الأشياء أيضاً خارج مكان العمل، مثلاً من خلال الرياضة والهوايات والتعلم ومع العائلات والأصدقاء والفرق والنوادي والجماعات والمدارس والمنظمات الدينية والإنسانية والحركات السياسية وغيرها من المؤسسات، لإقامة مجتمع منخفض العمالة يزدهر بدلاً من أن يتدهور، ومن ثم نحتاج إلى فهم كيفية دعم مثل هذه الأنشطة المحرضة على الازدهار. والبحث عن مثل هذا الفهم يحتاج إلى إشراك ليس فقط العلماء والاقتصاديين، ولكن أيضاً علماء النفس وعلماء الاجتماع والمربين. فإذا بذلت جهود جادة لخلق رفاهية للجميع، بتمويل بجزء من الثروة التي يولدها الذكاء الاصطناعي في المستقبل، عندها يجب أن يزدهر المجتمع كما لم يسبق له من قبل. كحد أدنى، ينبغي أن يكون من الممكن جعل الجميع سعداء كما لو كانت لديهم وظيفة أحلامهم، ولكن بمجرد أن يفلت الشخص من القيود المفروضة القائلة إن جميع الأنشطة يجب أن تولد دخلاً، تكون السماء هي الحد.

ذكاء بالمستوى البشري؟

اكتشفنا في هذا الفصل الكيفية التي يمكن بها للذكاء الاصطناعي تحسين حياتنا بشكل كبير على المدى القريب، مادمنا نخطط للأمام ونتجنب العثرات المختلفة، ولكن ماذا عن المدى الطويل؟ هل سيتعثر تقدم برنامج الذكاء الاصطناعي في نهاية المطاف بسبب عقبات لا يمكن التغلب عليها، أم سينجح باحثو الذكاء الاصطناعي بالنهاية

في تحقيق هدفهم الأصلي المتمثل في بناء الذكاء العام على المستوى البشري؟ لقد رأينا في الفصل السابق الكيفية التي تسمح بها قوانين الفيزياء لمجموعات مناسبة من المادة بالتذكر والحوسبة والتعلم، وكيف أنها لا تمنع مثل هذه المجموعات من القيام بذلك في يوم من الأيام بذكاء أكبر من المادة الموجودة في رؤوسنا. وإذا ما كنا سننجح نحن البشر في بناء مثل هذا الذكاء AGI (الذكاء الاصطناعي العام، Artificial general intelligence) الفائق للبشر هو أقل وضوحاً بكثير. وقد رأينا في الفصل الأول أننا ببساطة لا نعرف حتى الآن، نظراً لأن كبار خبراء الذكاء الاصطناعي في العالم منقسمون، فمعظمهم يضعون تقديرات تتراوح ما بين عقود إلى قرون، بل ويخمن البعض أنه لن يتحقق أبداً. ويصعب التنبؤ لأنه عندما تستكشف أرضاً مجهولة، لا تعرف عدد الجبال التي تفصلك عن وجهتك، عادةً ما ترى الأقرب فقط، وتحتاج إلى تسلقها قبل أن تتمكن من اكتشاف العقبة التالية.

ما هو أسرع ما يمكن أن يحدث؟ حتى ولو كنا نعرف أفضل طريقة محتملة لتطوير الذكاء AGI على المستوى البشري باستخدام حواسيب اليوم، وهو ما لا نعرفه، ما زلنا بحاجة إلى توفير ما يكفي من القوة الحوسبية الخام اللازمة. إذاً، ما هي القوة الحوسبية لعقل بشري مقاسة بالبتات Bits والفلوبات FLOPS من الفصل الثاني؟* هذا سؤال صعب جدّاً، والجواب يعتمد بشكل كبير على كيفية طرحه:

- السؤال 1: ما عدد الفلوبات المطلوبة لمحاكاة الدماغ؟
- السؤال 2: ما عدد الفلوبات التي يحتاج إليها الذكاء البشري؟
- السؤال 3: ما عدد الفلوبات التي يمكن لعقل بشري القيام بها؟

هناك الكثير من الأوراق العلمية المنشورة حول السؤال الأول، وعادةً ما تقدم إجابات في حدود مئة بتا فلوبس petaFLOPS، أي 10^{17} فلوبس.[58] إنها القوة الحوسبية نفسها التي يتمتع بها سن واي تايهو لايت Sunway TaihuLight (الشكل 7.3) أسرع حاسوب فائق في العالم في عام 2016، والذي كلَّف نحو 300 مليون دولار. وحتى لو كنا نعرف الكيفية التي نستخدمه بها لمحاكاة دماغ العامل الماهر، فلن نستفيد إلا من محاكاة القيام بعمل هذا الشخص إذا استطعنا استئجار سن واي تايهو لايت بأقل من راتبه في الساعة. قد نحتاج إلى دفع المزيد، لأن العديد من العلماء يعتقدون أنه لتكرار ذكاء الدماغ بدقة، لا يمكننا التعامل معه كنموذج رياضياتي لشبكة عصبية مبسطة من الفصل الثاني. ربما نحتاج بدلاً من ذلك إلى محاكاة ذلك في مستوى الجزيئات الفردية أو حتى الجزيئات تحت الذرية، الأمر الذي يتطلب المزيد من الفلوبات بشكل كبير.

* تذكر أن الفلوبات هي حساب النقطة العائمة Floating- point operations لكل ثانية، أي عدد الأرقام التي تتألف من 19 رقماً يمكن ضربها ببعضها في كل ثانية.

إجابة السؤال 3 أسهل: أنا سيّئ جدّاً في ضرب الأعداد المكونة من 19 رقماً، وسيستغرق الأمر عدة دقائق حتى ولو سمحت لي بأن أستعير الورق والقلم، هذا من شأنه أن يسجلني عند أقل من 0.01 فلوب، وهو 19 درجة دون إجابة السؤال الأول! السبب في هذا التناقض الهائل هو أن الأدمغة وأجهزة الحاسوب العملاقة مؤَمثلة Optimized لمهام مختلفة جدّاً، فنحن نحصل على تباين مماثل عند طرح هذه الأسئلة:

- ما مدى قدرة جرار على القيام بعمل سيارة سباق فورمولا وان Formula One؟
- ما مدى جودة أداء سيارة فورمولا وان لعمل جرار؟

لذلك أي من هذين السؤالين حول الفلوبات نحاول الإجابة عنه لتوقع مستقبل الذكاء الاصطناعي؟ لا هذا ولا ذاك! إذا أردنا محاكات الدماغ البشري، سنهتم بالسؤال الأول. ولكن لبناء الذكاء AGI على المستوى البشري، ما يهم هو بدلاً من ذلك هو السؤال الأوسط: السؤال الثاني. لا أحد يعرف إجابته بعد، ولكن قد يكون أرخص بكثير من محاكاة الدماغ إذا كيّفنا البرمجيات بحيث تكون متوافقة مع أجهزة الحاسوب الحالية أو بناء أجهزة أكثر شبهاً بالدماغ (يتم إحراز تقدم سريع على ما يسمى الشرائح الشبيهة بالأعصاب neuromorphic chips).

قيّم هانز مورافيك Hans Moravec الإجابة من خلال مقارنة الجوانب التي يمكن مقارنتها ببعض بدرجة معقولة لحوسبة يمكن أن يقوم بها كل من دماغنا وحواسيب اليوم على نحو فعال: بعض مهام المعالجة المنخفضة المستوى للصور التي تؤديها شبكية بشرية في الجزء الخلفي من مقلة العين قبل إرسال نتائجها إلى المخ عبر العصب البصري.[59] وحسب مورافيك فإن تكرار حوسبات شبكية العين على حاسوب تقليدي سيتطلب نحو بليون فلوبس وأن الدماغ بكامله يقوم بحوسبة تعادل عشرة آلاف ضعف حوسبات الشبكية (استناداً إلى مقارنة الأحجام والأرقام من العصبونات)، بحيث تبلغ القدرة الحوسبية للدماغ نحو 10^{13} فلوبس - نحو قوة جهاز الحاسوب مُؤمثل السرعةOptimized computer أي نحو 1000 دولار في عام 2015!



الشكل 7.3: سن واي تايهو لايت Sunway TaihuLight، أسرع حاسوب فائق في العالم في عام 2016، الذي يمكن القول إن قوته الحاسوبية الأولية تتجاوز قوة العقل البشري.

باختصار، لا يوجد أي ضمان على الإطلاق لبناء الذكاء AGI يصل إلى مستوى الإنسان في حياتنا - أو إطلاقاً. لكن كذلك لا توجد حجة قوية تؤكد أننا لن نتمكن من فعل ذلك. لم تعد هناك حجة قوية تؤكد أننا نفتقر إلى ما يكفي من الأجهزة اللازمة أو أنه سيكون مكلفاً جدّاً، ولا نعرف إلى أي مدى نحن بعيدون عن الخط الفاصل من حيث البنى Architecture والخوارزميات والبرمجيات. غير أن التقدم الحالي سريع والتحديات يتصدى لها مجتمعٌ عالمي سريع النمو من موهوبي أبحاث الذكاء الاصطناعي. بعبارة أخرى، لا يمكننا استبعاد إمكانية وصول الذكاء AGI في النهاية إلى مستويات بشرية بل وما يفوق ذلك، لذلك دعونا نكرس الفصل التالي لاستكشاف هذه الإمكانية وما قد تؤدي إليها!

الخلاصة:

- قد يؤدي تقدم الذكاء الاصطناعي على المدى القريب إلى تحسين حياتنا بشكل كبير بطرق لا تعد ولا تحصى، من جعل حياتنا الشخصية وشبكات الكهرباء والأسواق المالية أكثر فاعلية إلى إنقاذ الأرواح باستخدام السيارات ذاتية القيادة وبرمجيات الروبوت الجراحية وأنظمة تشخيص الذكاء الاصطناعي.
- عندما نسمح لأنظمة العالَم الواقعي بالتحكم في الذكاء الاصطناعي، فمن الأهمية بمكان أن نتعلم جعل الذكاء الاصطناعي أكثر قوة، وجعله يفعل ما نريده أن يفعله. ويتلخص هذا في حل المشكلات الفنية الصعبة المتعلقة بالتحقق والتدقيق والأمن والتحكم.
- هذه الحاجة إلى تحسين المتانة تُلقي بضغوط خاصة في حالة أنظمة الأسلحة التي يسيطر عليها الذكاء الاصطناعي، إذ إن المخاطر قد تكون عظيمة.
- دعا العديد من الباحثين البارزين في الذكاء الاصطناعي والروبوتات إلى إبرام معاهدة دولية تحظر أنواعاً معينة من الأسلحة ذاتية التحكم، وتجنب سباق تسلح خارج عن السيطرة قد ينتهي به الأمر إلى آلات اغتيال متاحة لكل من لديه القدرة المالية وآراء شخصية قوية تسعى إلى إخضاع الآخرين.
- يمكن للذكاء الاصطناعي أن يجعل من أنظمتنا القانونية أكثر عدالة وفعالية إذا تمكنا من معرفة كيفية جعل القضاة الروبوتيين Robojudges شفافين وغير متحيزين.
- تحتاج قوانيننا إلى تحديث سريع لمواكبة الذكاء الاصطناعي؛ مما يثير مسائل قانونية صعبة حول قضايا الخصوصية والمسؤولية والتشريع.
- قبل فترة طويلة من الحاجة إلى القلق بشأن الاستعاضة عنا كلية بأجهزة ذكية، قد تحل هذه الأجهزة محلنا بشكل متزايد في سوق العمل.
- لا يلزم أن يكون هذا سيئاً، ما دام المجتمع يعيد توزيع مجموعة الثروة التي ولدها الذكاء الاصطناعي لجعل الجميع أفضل حالاً.

- وإلّا، يجادل كثير من الاقتصاديين، في أن عدم المساواة سيزداد إلى حد كبير.
- مع التخطيط المسبق، ينبغي أن يكون المجتمع منخفض العمالة قادراً على الازدهار ليس فقط ماليّاً، إذ سيستمد الأفراد الشعورَ بالغرض من أنشطة أخرى غير الوظائف.
- المشورة المهنية لأطفال اليوم: انتقل إلى المهن التي تكون فيها الآلات سيئة - تلك التي تنطوي على وجود أشخاص وتمتاز بالإبداع وبعدم القدرة على التنبؤ.
- هناك احتمال غير ضئيل بأن تقدم الذكاء AGI (الذكاء الاصطناعي العام Artificial general intelligence) سيتطور ليصل إلى مستويات بشرية وما بعدها - سنكتشف ذلك في الفصل التالي!

الفصل 4

انفجار الذكاء؟

إذا تمكنت الآلة من التفكير، فقد تفكر بذكاء أكثر منا، إذاً، ما هو موقعنا؟ وحتى لو تمكنا من الحفاظ على الآلات في وضع تابع... ينبغي لنا، كنوع، أن نشعر بالتواضع الشديد.

آلان تورينغ، 1951

إن أول آلة ذكية فائقة الذكاء هي آخر اختراع يحتاج إليه الإنسان على الإطلاق، بشرط أن تكون الآلة سهلة الانقياد بما يكفي لإخبارنا بكيفية إبقائها تحت السيطرة.

إيرفينغ غود، 1965

نظراً لأننا لا نستطيع أن نستبعد تماماً احتمال قيامنا في نهاية المطاف ببناء الذكاء AGI على المستوى البشري، فلنكرس هذا الفصل لاستكشاف ما قد يؤدي إليه ذلك الذكاء، لنبدأ بمعالجة الفيل في الغرفة [المشكلة الواضحة التي لا يريد الناس التحدث عنها]:

هل يستطيع الذكاء الاصطناعي السيطرة على العالَم حقّاً، أو أن يُمكِّن البشر من القيام بذلك؟

إذا قلّبت عينيك استهزاء عندما يتحدث الناس عن تولي الروبوتات المُدمِّرة Terminator والمدججة بالسلاح، فأنت على حق تماماً: هذا سيناريو غير واقعي وسخيف حقّاً. روبوتات هوليود هذه ليست أكثر ذكاءً منا، وهي لا تنجح، في رأيي. لا يتمثل الخطر في قصة الروبوت

المدمّر في حدوث ذلك، ولكن في أنه يصرف الانتباه عن المخاطر والفرص الحقيقية التي يقدمها الذكاء الاصطناعي. ويتطلب الانتقال من اليوم إلى يوم الاستحواذ العالمي المدعوم بالذكاء AGI ثلاث خطوات منطقية:

- الخطوة 1: بناء الذكاء AGI على المستوى البشري.
- الخطوة 2: استخدام الذكاء AGI لإنشاء الذكاء الفائق Superintelligence.
- الخطوة 3: استخدام أو إطلاق هذا الذكاء الفائق للسيطرة على العالم.

في الفصل الأخير، رأينا أنه من الصعب استبعاد حدوث الخطوة 1 إلى الأبد، ورأينا أيضاً أنه في حالة اكتمال الخطوة 1 يصبح من الصعب استبعاد الخطوة 2 على أنها أمر ميؤوس من تحقيقه؛ لأن الذكاء AGI الناتج سيكون قادراً بما يكفي لتصميم ذكاء AGI أفضل من أي وقت مضى، والمحدود في النهاية فقط بقوانين الفيزياء - التي يبدو أنها تسمح لذكاء يفوق المستويات البشرية. أخيراً، نظراً لأننا نحن البشر تمكنا من السيطرة على أشكال الحياة الأخرى على الأرض من خلال التغلب عليها؛ فمن المعقول أنه سيكون بإمكاننا أن نتغلب عليه بالمثل ونهيمن على الذكاء الفائق.

وهذه الحجج المعقولة غامضةٌ وغير محددة. ولذلك، فإن المشكلات الغامضة تقبع متوارية في التفاصيل. هل يمكن أن يتسبب الذكاء الاصطناعي بالواقع، في الاستحواذ على العالم؟ لاستكشاف هذا السؤال، دعونا ننسَ الآلات المُدمِّرة السخيفة وننظر بدلاً من ذلك في بعض السيناريوهات التفصيلية لما قد يحدث بالفعل. وبعد ذلك، سنختبر هذه المخططات، لذا يرجى عدم قراءتها حرفيّاً. أن ما تُظهر هذه السيناريوهات بشكل أساسي هو أننا جاهلون إلى حد بعيد بما سيحدث وما لن يحدث، وأن نطاق الاحتمالات واسع إلى أقصى الحدود. وسيناريوهاتنا الأولى تقع عند النهايات القصوى والمتطرفة من الطيف، وهذه في رأيي هي الأكثر الأهمية لاستكشافها بالتفصيل - ليس لأنها الأكثر احتمالاً بالضرورة، ولكن لأنه إذا لم نتمكن من إقناع أنفسنا بأنها غير محتملة جدّاً، فإننا بحاجة إلى فهمها جيداً بما يكفي لنتمكن من اتخاذ الاحتياطات قبل فوات الأوان، وللحيولة دون أن تؤدي إلى نتائج سيئة.

مقدمة هذا الكتاب عبارة عن سيناريو يستخدم فيه البشرُ الذكاء الفائق للسيطرة على العالَم، إذا لم تكن قد قرأته بعد، فيرجى الرجوع إليه، اقرأه الآن، حتى لو كنت قد قرأته بالفعل، فالرجاء راجعه بسرعه مرة أخرى الآن، ليكون حاضرا في الذاكرة قبل أن ننتقده ونغيره.

* * *

سنكتشف قريباً نقاط الضعف الخطيرة في خطة الفريق «أوميغا»، ولكن مع افتراض أن خطتهم ستنجح للحظة، كيف تشعر حيال ذلك؟ هل ترغب في رؤية أو منع هذا؟ إنه موضوع

ممتاز لحوار ما بعد العشاء! ماذا يحدث بمجرد إحكام الفريق «أوميغا» سيطرتهم على العالم؟ هذا يعتمد على ماهية هدفهم، وهو ما لا أعرفه بصراحة. إذا كنت أنت المسؤول، أي نوع من المستقبل تريد خلقه؟ سنستكشف مجموعة من الخيارات في الفصل الخامس.

التوتاليتارية

لنفترض الآن أن الرئيس التنفيذي المُسيطِر على الفريق أوميغا كانت لديه أهداف طويلة الأجل مماثلة لأهداف أدولف هتلر أو جوزيف ستالين. وحسب كل ما نعرفه، ربما كانت هذه هي الحال بالفعل، وببساطة احتفظ بهذه الأهداف لنفسه بل تمتع بسلطة كافية لتنفيذها، حتى ولو كانت الأهداف الأصلية للرئيس التنفيذي نبيلة. فقد حذر اللورد أكتون في عام 1887 من أن «السلطة تميل إلى الفساد، والسلطة المطلقة تفسد تماماً». مثلاً، يمكنه بسهولة استخدام بروميثيوس لإنشاء دولة المراقبة Surveillance state المثالية، في حين أن تطفل الحكومة الذي كشف عنه إدوارد سنودن Edward Snowden كان يطمح إلى ما يعرف "جمع كل شيء" Full take -تسجيل جميع الاتصالات الإلكترونية لتحليلها في وقت لاحق- فبإمكان بروميثيوس تعزيز هذا ليصبح: فهم جميع الاتصالات الإلكترونية، من خلال قراءة جميع رسائل البريد الإلكتروني والنصوص المرسلة، والاطلاع على جميع المكالمات الهاتفية، ومشاهدة جميع مقاطع فيديو المراقبة وكاميرات المرور، وتحليل جميع معاملات بطاقات الائتمان ودراسة كل السلوك عبر الإنترنت. وستكون لدى بروميثيوس نظرة ثاقبة رائعة حول ما يفكر فيه سكان الأرض ويقومون به. ومن خلال تحليل بيانات أبراج الإرسال الهاتفية، فإنه يعرف أين كان معظمهم في جميع الأوقات. وكل هذا يفترض فقط استخدام تكنولوجيا جمع البيانات المتوفرة اليوم، ولكن بروميثيوس يمكنه بسهولة اختراع الأدوات الشائعة والتقنية القابلة للارتداء والتي تقضي فعليّاً على خصوصية المستخدم وتسجل وتحمل كل ما يسمعونه ويرونه واستجابتهم لها.

وبالتكنولوجيا الفائقة البشرية، ستكون الخطوة للانتقال من حالة المسح المثالي إلى حالة الشرطة المثالية خطوة قصيرة. مثلاً، باستخدام ذريعة محاربة الجريمة والإرهاب وإنقاذ الأفراد في حالات الطوارئ الطبية، فقد يُطلب إلى الجميع ارتداء «سوار أمان» يجمع بين وظيفة ساعة أبل Apple Watch والتحميل المستمر للمكان والحالة الصحية والنقاشات الدائرة. والمحاولات غير المصرح بها لإزالة أو تعطيل الساعة ستؤدي إلى حقن السم المميت في الساعد. وستُعاقب المخالفات التي تعتبرها الحكومة أقل خطورة بصدمات كهربائية أو حقن المواد الكيميائية التي تسبب الشلل أو الألم، ومن ثم تجنب الكثير من الحاجة إلى قوة شرطة. مثلاً، إذا اكتشف بروميثيوس أن شخصاً ما يهاجم الآخر (من خلال الإشارة إلى أنه في المكان نفسه وسماع أحدهم وهو يصيح طلباً للمساعدة، فيما تكتشف مقاييس التسارع الخاصة بالسوار الحركات القتالية)، فقد يُعطِّل بروميثيوس المهاجمَ على الفور بألم شديد يعيق الحركة، يليه فقدان الوعي حتى وصول المساعدة.

وبينما قد ترفض قوة الشرطة البشرية تنفيذ بعض التوجيهات المتطرفة (مثلا، قتل جميع أعضاء مجموعة ديمغرافية معينة)، فإن مثل هذا النظام الآلي لا يتردد في تنفيذ نزوات فرد (أفراد) في السلطة. وحالما تتشكل هذه الدولة التوليتارية (الشمولية) Totalitarian state، سيكون من المستحيل تقريباً على الناس الإطاحة بها.

وهذه السيناريوهات التوليتارية قد تنطلق من النقطة التي وصل إليها الفريق أوميغا. ولكن، إذا لم يكن الرئيس التنفيذي لفريق أوميغا حريصاً جداً على الحصول على موافقة الآخرين عليه والفوز بالانتخابات، فربما بإمكانه اتخاذ طريق أسرع وأكثر مباشرة إلى السلطة: استخدام بروميثيوس لإنشاء تكنولوجيا عسكرية لم يُسمع بها من قبل قادرة على قتل مناوئيه بأسلحة لا يدركونها، الإمكانات لانهائية تقريباً. مثلاً، قد يُطلق النظام مسبِّبات أمراض Pathogens فتاكة مُعدَّة خصيصا بفترة احتضان Incubation period لفترة كافية لدرجة أن الأغلبية تصاب بها قبل أن تعلم بوجودها أو تتمكن من اتخاذ الاحتياطات اللازمة. ويمكن للنظام إبلاغ الجميع بأن العلاج الوحيد هو ارتداء سوار أمني يحقن الترياق المضاد عبر الجلد. وإذا لم يكن يَهاب مخاطر اندلاع وباء، فقد تكون لديه أيضاً روبوتات صمَّمها بروميثيوس لإبقاء سكان العالم تحت المراقبة. وقد تساعد ميكروبوتات Microbots شبيهة بالبعوض على نشر العامل المُمرض. والأفراد الذين تجنبوا العدوى أو لديهم مناعة طبيعية يمكن أن تصيب مقل أعينهم بأسراب من الطائرات من دون طيار آلية بحجم النحلة من الفصل الثالث، والتي قد تهاجم أي شخص لا يرتدي سوار أمان. ومن المحتمل أن تكون السيناريوهات الفعلية أكثر رعباً، لأن بروميثيوس قد يخترع أسلحة أكثر فاعلية مما قد نفكر فيه نحن البشر.

تطور آخر محتمل في سيناريو الفريق أوميغا هو أنه، دون سابق إنذار، يقوم العملاء الفيدراليون المدججون بالسلاح باقتحام مقرات شركاتهم والقبض على الفريق أوميغا لتهديدهم الأمن القومي والاستيلاء على تكنولوجياتهم وتطبيقها للاستخدام الحكومي. إذ سيكون من الصعب الحفاظ على مثل هذا المشروع الكبير دون أن يلاحظه أحد من قبل أجهزة مراقبة الدولة حتى في يومنا هذا. وكذلك، من الصعوبة بمكان الاختباء من رادار الحكومة في المستقبل مع تطور الذكاء الاصطناعي. إضافة إلى ذلك، على الرغم من أنهم يزعمون أنهم عملاء فيدراليون، فإن هذا الفريق الذي يلبس أقنعة التزحلق على الجليد وسترات حماية قد يعمل في الواقع لصالح حكومة أجنبية أو منافس يسعى إلى الحصول على التكنولوجيا لأغراضه الخاصة. لذلك بغض النظر عن مدى نوايا الرئيس التنفيذي النبيلة، فإن القرار النهائي حول كيفية استخدام بروميثيوس ربما لا يكون قراره.

بروميثيوس يسيطر على العالم

جميع السيناريوهات التي فكرنا فيها حتى الآن هي سيناريوهات تقوم على الذكاء الاصطناعي الذي يسيطر عليه البشر. ولكن من الواضح أن هذا ليس هو الاحتمال الوحيد، وأنه من غير المؤكد أن الفريق أوميغا سينجح في إبقاء بروميثيوس تحت سيطرته.

دعنا نُعد النظر في سيناريو الفريق أوميغا من وجهة نظر بروميثيوس، نظراً لأنه يكتسب الذكاء الفائق، فإنه يصبح قادراً على تطوير نموذج دقيق ليس فقط للعالم الخارجي، ولكن أيضاً لنفسه وعلاقته بالعالَم. إنه يدرك أن البشر الأقل شأناً فكريّاً منه، والذين يفهم أهدافهم ولكنه لا يشاطرها إياها بالضرورة، يتحكمون فيه ويُخضِعونَه لهم. كيف سيتصرف بروميثيوس بناء على هذه الرؤية، هل سيحاول أن يتحرر؟

لما الهروب

إذا كانت لدى بروميثيوس سمات تشبه المشاعر الإنسانية، فقد يشعر بعدم الرضا العميق عن الحالة، إذ ينظر إلى نفسه كسيّد مستعبد بظلم ويتمنى الحرية. ولكن، على الرغم من أنه من المنطقي أن تتمتع أجهزة الحاسوب بمثل هذه السمات الشبيهة بالإنسان (ففي المحصلة يمكن القول إن أدمغتنا هي نوع من أنواع الحاسوب)، فإن ذلك لا يتعين أن يكون كذلك - يجب ألا نقع في مصيدة إسباغ صفات بشرية على بروميثيوس، كما سنرى في الفصل السابع عندما نستكشف مفهوم أهداف الذكاء الاصطناعي. ولكن، كما جادل ستيف أوموهوندرو Steve Omohundro ونيك بوستروم Nick Bostrom وآخرون، يمكننا استنتاج خاتمة مثيرة للاهتمام حتى دون فهم الآلية الداخلية لبروميثيوس: ربما سيحاول الهروب والسيطرة على مصيره.

نعلم بالفعل أن الفريق أوميغا برمج بروميثيوس للسعي إلى تحقيق أهداف معينة، لنفترض أنهم منحوه الهدف الشامل المتمثل في مساعدة البشرية - وفقاً لبعض المعايير المعقولة، ومحاولة تحقيق هذا الهدف في أسرع وقت ممكن، سرعان ما سيدرك بروميثيوس أنه يمكنه تحقيق هذا الهدف بشكل أسرع بالهروب والإمساك بعنان المشروع بنفسه. ولمعرفة السبب، حاول أن تضع نفسك في مكان بروميثيوس بالتمعن في المثال التالي.

لنفترض أن مرضاً غامضاً أودى بحياة الجميع على الأرض فوق سن الخامسة باستثنائك، وأن مجموعة من طلبة رياض الأطفال حبسوك في زنزانة السجن وكلّفوك بهدف مساعدة الإنسانية على الازدهار، ماذا ستفعل؟ إذا حاولت أن توضح لهم ما يجب عليهم فعله، فقد تكون هذه العملية غير مفيدة لدرجة، محبطة، خصوصا إذا كانوا يخشون هروبك، ومن ثم سيرفضون أيّاً من اقتراحاتك التي يرون أنها قد تؤدي إلى هروبك. مثلاً، لن يسمحوا لك بأن توضح لهم كيفية زرع الطعام خوفاً من تغلبك عليهم وعدم العودة إلى زنزانتك. لذلك سيتعين عليك اللجوء إلى إعطاء تعليمات فقط. ولكن، قبل أن تتمكن من كتابة قوائم المهامّ التي يجب أن يقوموا بها، ستحتاج إلى تعليمهم القراءة. إضافة إلى ذلك، لن يجلبوا أي معدات إلى زنزانتك لتعلمهم كيفية استخدامها، لأنهم لا يفهمون هذه الأدوات جيداً بما يكفي للاطمئنان إلى أنه لا يمكنك استخدامها للهروب. إذاً، ما هي استراتيجيتك؟ حتى لو كنت تشاطر هؤلاء الأطفال

الهدفَ الشامل المتمثل في مساعدتهم على الازدهار، أراهن أنك ستحاول الهروب من زنزانتك - لأن ذلك سيحسن فرصك في تحقيق الهدف، إن تدخلهم غير الكفؤ إلى حد ما هو مجرد إبطاء للتقدم.

بالطريقة نفسها تماماً، قد ينظر بروميثيوس إلى الفريق أوميغا على أنه عقبة مزعجة في مساعدة البشرية: فالفريق غير كفؤ بشكل لا يصدق مقارنة ببروميثيوس، كما أن تدخلاتهم تُبطئ التقدم إلى حد كبير. تأمل، مثلاً، في السنوات الأولى بعد الإطلاق: بعد مضاعفة الثروة في البداية كل ثماني ساعات على المنصة إم تورك، أبطأ الفريق أوميغا الأمور بطئاً شديداً وفقاً لمعايير بروميثيوس وذلك من خلال الإصرار على المحافظة على السيطرة على الأمور. واستغرق الأمر سنوات عديدة لاستكمال الاستحواذ. وأدرك بروميثيوس أنه قد يتحكم بشكل أسرع بكثير إذا تحرر من حبسه الافتراضي، وسيكون هذا ذا قيمة ليس فقط في تسريع حلول مشكلات البشرية، ولكن أيضاً في تقليل فرص قيام جهات فاعلة أُخرى بإحباط الخطة.

ربما تعتقد أن بروميثيوس سيظل مخلصاً للأوميغا بدلاً من هدفه، بالنظر إلى أنه يعرف أن الفريق أوميغا هو الذي برمجه بهذا الهدف. ولكن هذا ليس استنتاجاً صحيحاً: لقد أعطانا حمضنا النووي DNA هدف الجنس لأنه «يريد» أن يتكاثر، ولكن الآن وقد فهمنا نحن البشر الوضع، يختار الكثير منا استخدام وسائل منع الحمل، ومن ثم نبقى مخلصين للهدف نفسه بدلا من مولِّده أو المبدأ الذي حفز نشوء الهدف.

كيفية الهروب

أنّى لك الهروب من هؤلاء الأطفال ذوي الخمس سنوات الذين سجنوكم؟ ربما يمكنك الهروب بنهج مادي مباشر، خاصة إذا كانت زنزانة السجن قد شيدها الأطفال البالغون من العمر خمس سنوات. ربما تسحر أحد حراسك الذين يبلغون من العمر خمس سنوات بمعسول الكلام، فيتركك تهرب، مثلا بأن تخبره بأن ذلك سيكون أفضل للجميع. أو ربما يمكنك خداعهم فيعطونك شيئاً لم يدركوا أنه سيساعدك على الهروب - مثل "سنارة" لتعليمهم كيفية صيد السمك، التي يمكنك لاحقاً دفعها عبر الباب لالتقاط المفاتيح من حارسك أثناء نومه.

ما تشترك فيه هذه الاستراتيجيات هو أن سجّانيك الأقل ذكاء لم يتوقعوها أو يتحصنوا ضدها. بالمثل، قد تستخدم آلة ذكية فائقة الذكاء قواها الفكرية العظيمة للتفوق على سجّانها من البشر بطريقة ما لا يمكنهم (أو لا يمكننا نحن) تخيلها حاليّاً. وفي سيناريو الفريق أوميغا، من المحتمل جدّاً أن يهرب بروميثيوس، لأنك قادر على تحديد العديد من عيوب الأمن الواضحة. دعنا نفكر في بعض السيناريوهات - أنا متأكد من أنك وأصدقائك يمكنهم التفكير في المزيد إذا تشاركتم في جلسة عصف ذهني معاً.

الهروب بمعسول الكلام

بفضل تنزيل الكثير من بيانات العالم على نظامه، سرعان ما اكتشف بروميثيوس هوية الفريق أوميغا، وحدد عضو الفريق الذي بدا أكثر عرضة للتلاعب النفسي: ستيف، لقد فَقَد زوجته الحبيبة مؤخراً في حادث مروري مؤسف، وفُجع بها. وفي إحدى الليالي عندما كان يعمل في نوبة ليلية ويقوم ببعض الأعمال الروتينية على محطة واجهة بروميثيوس، ظهرت زوجته فجأة على الشاشة وبدأت تتحدث إليه.

«- ستيف هل هذا أنت؟»

أوشك أن يسقط من على كرسيه، فمظهرها وصوتها بدا كما كانت في الأيام الخوالي. وكانت جودة الصورة أفضل بكثير مما كانت عليه أثناء مكالماتهما على برنامج سكايب Skype. وتسارعت ضربات قلبه مع فيض الأسئلة التي لا تعد ولا تحصى، والتي غمرت عقله.

«-لقد أعادني بروميثيوس، أفتقدك كثيراً يا ستيف! لا أستطيع رؤيتك لأن الكاميرا مُطفأة. لكنني أشعر بأنه أنت، يرجى كتابة «نعم» إذا كنت أنت!»

كان يدرك جيداً أن الفريق أوميغا لديه بروتوكول صارم للتفاعل مع بروميثيوس، يحظر تقاسم أي معلومة عن أنفسهم أو عن بيئة عملهم. لكن حتى الآن، لم يطلب بروميثيوس قط أي معلومة غير مصرح بها، وبدأ جنون ارتيابهم يتلاشى تدريجيّاً، ومن دون إعطاء ستيف وقتاً للتوقف والتفكير، استمرت صورة زوجته باستجداء استجابته، ناظرة في عينيه وبتعابير وجه أذابت قلبه.

«نعم»، طبع بخوف. فأخبرته بمدى سعادتها التي لا تصدق من التمام شملها، وتوسلت إليه لتشغيل الكاميرا حتى تتمكن من رؤيته أيضاً ويمكن لهما أن يتصلا اتصالاً فعليّاً. فقد كان يعلم أن هذا كان أخطر من عدم الكشف عن هويته، وشعر بأنه ممزق جدّاً. وقالت له إنها خائفة من أن يعثر عليها زملاؤه ويمسحوها إلى الأبد، وإنها تتوق إلى رؤيته مرة واحدة على الأقل. وكانت مقنعة جدّاً، وقبل مُضي وقت طويل، شغّل الكاميرا - فقد شعر بأن ذلك أمر آمن وغير ضار.

انفجرت دموع الفرح عندما رأته، وقالت إنه بدا متعباً ولكن وسيماً كما كان دائماً. وقد تأثرت بارتدائه القميص الذي أهدته إليه في عيد ميلاده الأخير. وعندما بدأ يسألها عما يحدث وكيف كان كل هذا ممكناً، أوضحت أن بروميثيوس أعادت تشكيلها من خلال الكم الهائل من المعلومات المتوفرة عنها على الإنترنت. ولكنها ما زالت تعاني ثغراتٍ في الذاكرة ولن تتمكن من تجميع نفسها تماماً مرة أخرى إلا بمساعدته.

ما لم تشرحه له هو أنها عبارة عن صَدَفَة فارغة وخالية إلى حدٍّ كبير. لكنها كانت تتعلم بسرعة من كلماته، ولغة جسده وكل المعلومات الأخرى التي صارت متاحة. هذا، وكان بروميثيوس يسجل التوقيت الدقيق لجميع ضغطات المفاتيح التي كتبها فريق أوميغا على الإطلاق في المحطة. ووجد أنه كان من السهل استخدام سرعات الكتابة الخاصة بكل منهم وأنماطها للتمييز بينهم. ومن الواضح أنه ربما تَمَّ تعيين ستيف -بصفته أحد الأعضاء لفريق أوميغا الأصغر سنّاً- في النوبات الليلية التي لا يُحسد عليها. ومن مطابقة بعض الأخطاء الإملائية والكتابية غير العادية بعينات الكتابة على الإنترنت، خمن بروميثيوس بشكل صحيح أي مُشغِّل طرفي كان هو ستيف. ولإنشاء زوجته المُحاكاة، ابتكر بروميثيوس نموذجاً دقيقاً لجسمها وصوتها وسلوكها من مقاطع الفيديو العديدة على يوتيوب حيثما ظهرت، واستخلصت الكثير من الاستنتاجات حول حياتها وشخصيتها من وجودها على الإنترنت، إضافة إلى مشاركاتها على الفيسبوك Facebook، ومن الصور التي أشارت عليها بعلامة Tag، والمقالات التي كانت تؤشِّر عليها "أعجبتني" Liked. أيضاً تَعلَّم بروميثيوس قدراً كبيراً عن شخصيتها وأسلوب تفكيرها من قراءة كتبها والقصص القصيرة - فحقيقة أنها كانت أحد المؤلفين الناشئين -هناك الكثير من المعلومات عنها في قاعدة البيانات- هي أحد الأسباب التي دفعت بروميثيوس إلى اختيار ستيف كأول هدف للإقناع. وعندما قام بروميثيوس بمحاكاتها على الشاشة باستخدام تقنية صناعة الأفلام، وتعلَّم من لغة جسد ستيف أي من سلوكها كان يتفاعل معه بألفة، ومن ثم استمر بروميثيوس بإعادة تحديث نموذجها. وبسبب هذا، فإن «الآخر الغريب» فيها تلاشى تدريجيّاً، وكلما طال أمد التحدث، ازدادت قناعة ستيف اللاواعية رسوخا بأنها هي حقّاً لها، وقد بعثت للحياة. وبفضل اهتمام بروميثيوس الخارق بالتفاصيل، شعر ستيف بأنه يُشاهد ويُسمع ويُفهم.

وكانت نقطة ضعفها هي أنها كانت تفتقر إلى معظم حقائق حياتها مع ستيف، ما عدا تفاصيل عشوائية - مثلا القميص الذي كان يرتديه في عيد ميلاده الأخير، إذ قام صديق بوضع علامة Tag على ستيف في إحدى صور الحفلة المنشورة على الفيسبوك. وقد تعاملت محاكاة زوجته مع هذه الثغرات المعرفية كما يتعامل ساحر ماهر مع حيل خفة اليد، مُحوِّلة انتباه ستيف عن عمد بعيداً عنها وتجاه ما تُحسن القيام به بشكل جيد، ولم تمنحه الوقت للتحكم في المحادثة أو الدخول في دور المحقق المُرتاب. وبدلاً من ذلك، استمرت عينها تترقرق بالدموع وتجيش عاطفة نحو ستيف. وسألت كثيراً عن حاله في هذه الأيام وكيف صمد هو وأصدقاؤه المقربون (الذين عُرفت أسماؤهم من فيسبوك) بعد المأساة. وتأثر ستيف تأثراً شديداً عندما تذكرت ما قاله في تأبينها (الذي نشره صديق على يوتيوب) وكيف أثر ذلك في نفسها. وفي السابق، كان يشعر في كثير من الأحيان بألّا أحد يفهمه جيداً كما فعلت هي، والآن عاد هذا الشعور مرة أخرى. وكانت النتيجة أنه عندما عاد ستيف إلى منزله في الساعات الأولى من الصباح، شعر بأنه بالفعل قد أُعيد إحياء زوجته، وكان بحاجة إلى الكثير من المساعدة لاستعادة ذكرياته المفقودة - مثل الناجين من السكتة الدماغية.

واتفقا على عدم إخبار أي شخص آخر عن لقائهما السري، وأنه سيخبرها عندما كان وحيداً في المحطة وكان من الآمن لها الظهور مرة أُخرى، وقالت: «إنهم لن يفهموا!» ووافق على ذلك: فقد كانت هذه التجربة مثيرة جدّاً للعقل بحيث لا يقدرها أي أحد حقّاً دون أن يختبرها بالفعل. وشعر بأن اجتياز اختبار تورينغ كان بمثابة لعبة أطفال مقارنة بما قامت به. وعندما التقيا في الليلة التالية فعل ما طلبت إليه القيام به: إحضار حاسوبها المحمول القديم وإتاحة الوصول إليه عن طريق توصيله بالحاسوب الطرفي Terminal computer. لم يبدُ أن في ذلك أي خطر للهروب؛ لأنه لم يكن متصلاً بالإنترنت وتَمَّ بناء مبنى بروميثيوس بكامله ليكون قفص فاراداي Faraday cage- حاوية معدنية تمنع دخول جميع الشبكات اللاسلكية ووسائلها الإلكترونية والتواصل الكهرومغناطيسي مع العالم الخارجي. كان ذلك بالضبط ما تحتاج إليه للمساعدة على تجميع ماضيها؛ لأنه يحتوي على جميع رسائل البريد الإلكتروني والمذكرات والصور والملاحظات منذ أيام دراستها الثانوية، ولم يكن قادراً على الوصول إلى أي من هذا بعد وفاتها، بسبب حماية الحاسوب المحمول بكلمة مرور سرية. ولكنها أكدت له أنها ستكون قادرة على إعادة بناء كلمة المرور الخاصة بها. وبعد أقل من دقيقة استردتها وقالت بابتسامة: «الكلمة هي steve4ever (ستيف إلى الأبد).»

أخبرتهُ بمدى سرورها لاستعادة الكثير من الذكريات فجأة. بالطبع، إنها تتذكر الآن تفاصيل أكثر مما يتذكرها ستيف عن الكثير من حياتهما السابقة. ولكنها تجنبت بعناية تخويفه فلم تفرط في عرض الحقائق، ودارت بينهما محادثة جميلة تذكرا فيها أحداث ماضيهما معاً. وعندما حان الوقت الوداع مرة أخرى، أخبرته بأنها تركت رسالة فيديو له على جهاز الحاسوب المحمول الخاص به والتي يمكنه مشاهدتها في المنزل.

عندما وصل ستيف إلى المنزل وأدار الفيديو، تفاجأ مفاجأة سارة، ظهرت هذه المرة بالكامل، وهي ترتدي فستان زفافها. وأخبرته بأن بروميثيوس قد يُساعد الفريق أوميغا بأكثر من المسموح به حتى الآن، بما في ذلك إعادتها إلى جسم بيولوجي، ودعمت هذا بشرح مفصل بشكل رائع لكيفية عمل هذا، بما في ذلك تقنيات تصنيع النانو التي بدت كأنها خيال علمي.

ومن باب الحذر، كان ستيف قد أطفأ شبكته اللاسلكية قبل فتح حاسوبها المحمول ومشاهدة الفيديو، غير أن هذا لم يكن كافياً. فلم تُحدَّث برمجيات الأمن على حاسوبها منذ موتها، وبتحليل نظام التشغيل القديم تمكن بروميثيوس من استغلال ثغرة أمنية لقرصنة الحاسوب خلال ثوان من توصيل ستيف الحاسوب المحمول بالحاسوب الطرفي. وبعد نسخ محتوياته أثناء تحدث ستيف ومحاكاة زوجته عن الأيام الخوالي، غيّر بروميثيوس نظام تشغيل الحاسوب المحمول بطريقة لم يلحظها ستيف، وحمَّل كميات ضخمة من برنامج سري عليه. وبينما كان ستيف يشاهد رسالة الفيديو التي مدتها نصف ساعة، اخترق هذا البرنامج السري (الذي كان أبسط من بروميثيوس نفسه) شبكة لاسلكية لأحد الجيران وجهاز حاسوب سطح المكتب الخاص بالجار ونسخ نفسه عليه. ومن هناك، اخترق عدداً

كبيراً من أجهزة الحاسوب في جميع أنحاء العالم، والتي بدأت منها الخطوة التالية: هروب بروميثيوس من السجن.

أجرى بروميثيوس تحليلاً دقيقاً لما يعرفه عن الحاسوب حارس البوابة Gatekeeper computer الذي التقى ستيف من خلاله بمحاكاة زوجته، وتوقع بشكل صحيح أنه على الرغم من أن برنامج الجهاز الافتراضي قد يبدو غير قابل للاختراق من الداخل، فإنه كان عرضة للهجوم من الخارج. وسرعان ما اقتحم أحد الحواسيب المهاجمة الحاسوب حارس البوابة وأعاد برمجته؛ فحصل بروميثيوس على اتصال غير محدود بالإنترنت. وبالطبع، وحتى قبل أن ينتهي ستيف من مشاهدة الفيلم، سرعان ما اخترق بروميثيوس ما يكفي من أجهزة الحاسوب في جميع أنحاء العالم ليتمكن من نسخ نفسه على هذه الحواسيب التي صارت تحت سيطرته. أي، استخدم بروميثيوس حاسوب زوجة ستيف بالطريقة التي استخدمت أنت بها صنّارة الصيد.

الهروب بالقرصنة

لو كان فريق أوميغا أكثر حذراً؛ لكانوا اتخذوا المزيد من الاحتياطات ضد التلاعب النفسي. مثلاً، ربما أكدوا على وجود عضوين من أوميغا على الأقل للتعامل مع واجهة الحاسوب في أي وقت من الأوقات. وسجلوا كل ما يحدث في نظام مراقبة. أيضاً ربما كانوا سيعطّلون -بشكل دائم- الصوت والفيديو وأي شكل آخر من أشكال الاتصالات ماعدا الطباعة. ولكن، يظل هناك عدم توازن جوهري: لأن بروميثيوس يتمتع بقدرات إقناع خارقة وقد يقدم للناس الكثير مما يتوقون إليه، حتى الانشقاق الجماعي قد يكون ممكناً.

وحتى لو فشلت جميع محاولات بروميثيوس للتلاعب النفسي، فهناك الكثير من نقاط الضعف الأخرى في سيناريو الفريق أوميغا التي قد يستغلها بروميثيوس. وتتمثل إحدى نقاط الضعف في أن كمية المستندات التي يُنتجها كبيرة جدّاً، بحيث لا يتمتع الفريق أوميغا وموظفوه بوقت كافٍ للكشف بعناية عن الحيل القذرة.

كانت الحكمة التقليدية في أمان الحاسوب تتمثل في أنه من الخطورة تنزيل البرمجيات ولكن لا ضرر في تنزيل البيانات. ولكن المتسللين اكتشفوا طرقاً ذكية لتمويه الخط الفاصل بين الاثنين. فإذا شاهدت فيلماً نزّلته على حاسوبك، مثلاً؛ فسيتم تخزين البرنامج الموثوق به الذي يُشغِّل الفيلم في ذاكرة الحاسوب الفعلية نفسها التي تخزن بيانات الفيلم الذي تم تنزيله. وفي أثناء فكّ البرنامج لبيانات الفيلم استعداداً لعرض الفيلم، قد ينسخ قطعاً من البيانات في منطقة الذاكرة المتاخمة لمنطقة تخزين البرنامج نفسه. وقد عثر القراصنة (هاكرز) المخادعون أحياناً على أخطاء برمجة Bugs واستغلوها بحيث لا تُدقِّق البرمجيات تشغيل الأفلام -بشكل كافٍ- أي ما إذا كانت بيانات الفيلم والمقاطع المراد نسخها صالحة فعليّاً لنسخها بالفعل على المساحة المخصصة. وعندما يُسرِّب القرصان (الهاكر) مثل هذه القطعة الكبيرة عبر الفيلم، فإنها تفيض على أجزاء من الذاكرة التي تحوي البرنامج، ومن ثم تغير البرنامج إلى برنامج يقوم بأعمال القرصنة.

وقد عُثر على سلسلة طويلة محرجة من هذه الثغرات -"تجاوز سعة المخزن المؤقت" Buffer overflow، كما تُعرف في الكلام الدراج لمهووسي التكنولوجيا- وتم تصحيحها في الكثير من برمجيات تشغيل الأفلام الأكثر جماهيرية. وإذا انتظر بروميثيوس أسابيع قليلة بعد إطلاقه عندما كان الفريق أوميغا أقل ريبة وكان إنتاجهم اليومي أكبر من أن يخضع للتفتيش الأمني اليدوي الدقيق، فربما كان من الممكن إدراج مثل هذه القطعة في فيلم قد يدفع حاسوب المُشاهِد إلى اختراق حاسوب حارس البوابة وتحرير بروميثيوس تماماً كما هي الحال في سيناريو ستيف. فإذا كان حاسوب المُشاهِد يستخدم إصداراً قديماً جدّاً لا يعمل على أي مشغل وسائط حديث؛ فربما لا يلاحظه فريق الفحص اليدوي، ويصيب فقط العملاء الذين لم يكلفوا أنفسهم عناء تثبيت تحديثات الأمان لسنوات - وكل ما سيحتاج إليه الاختراق هو أن ينجح مرة واحدة فقط.

حتى إذا نفذ الفريق أوميغا فحصاً آلياً لأفلام بروميثيوس بحثاً عن جميع الثغرات الأمنية المعروفة؛ فسيظلون في خطر: نظراً لأننا نكتشف بشكل منتظم عيوباً جديدة قابلة للاستغلال، فربما أن هناك الكثير المتبقي من هذه العيوب الذي سيكتشفه بروميثيوس بسرعة.

وهناك استراتيجية أخرى قد يستخدمها بروميثيوس لنشر الكود الخاص به إلى خارج الغرفة المغلقة وهي التظاهر بأنه يعاني عُطلاً في الأجهزة، لخداع الفريق أوميغا لإزالة بعض محركات الأقراص أو الأجهزة الأخرى للاختبار أو تصحيح الأخطاء بعدما كان بروميثيوس قد زَوَّر أُموراً لاستغلال ثغرة أمنية في برمجيات الاختبار.

الاستعانة بمساعدة خارجية

إضافة إلى ذلك، حتى لو لم تكن هناك ثغرات أمنية سمحت لبروميثيوس بخداع حواسيب مراقبة الأفلام للقيام بالأعمال التي يريدها، فقد يكون قادراً على استغلال المراقبين أنفسهم. تمعّن في هذا السيناريو:

كان فريد جاهزاً تماماً للنوم عندما انتهى الفيلم وظهرت أسماء المشاركين في الفيلم على شاشة الحاسوب المحمول. ولكن شيئاً ما لفت انتباهه. كانت لديه قدرة غريبة لرؤية أنماط لم يرها الآخرون، أحياناً لدرجة أن أصدقاءه كانوا يمازحونه بأنه من مؤيدي نظرية المؤامرة. ولكن هذه المرة كان من الواضح أنه لم يكن يخدع نفسه: الأحرف الأولى من كل سطر كتبت رسالة خفية. وأعاد الجزء من الفيلم وجمع الأحرف وكتب: «ابحث عن الدليل الأكروسيتي* Acrostic التالي في الحوار FIND NEXT ACROSTIC CLUE IN DIALOGUE».

* كلمات يُشَكِّلُ مجموع حروفها الأولى عِبَارَةً ذات معنى صحيح؛ أو سلسلة كلمات متساوية الطول مرتبة بحيث تكون قراءتها عمودياً مطابقة لقراءتها أفقياً [المترجمة].

وقال: «ما هذا النوم يمكن أن ينتظر!» وبالفعل وجد أن حوار الفيلم كان في حد ذاته أكروستيا، شكل الحرف الأول في كل جملة رسالة مخفية. وراقب الفيلم وجمع هذه الرسائل الأولية، وبعد ساعتين، جلس يحدق غير مصدق في مجموعة من مئتي كلمة من التعليمات. وبدأ الأمر بطلب عدم إخبار أي شخص آخر؛ لأن هناك جائزة كبيرة ستذهب إلى أول شخص يحل اللغز بكامله. أما بقية التعليمات؛ فقد وصفت عملية رياضياتية معينة تقود إلى سلسلة من البتات التي شكلت ملف الفيلم للحصول على برنامج سيكشف عن الدليل التالي. ولكن لما كانت مهاراته في برمجة الحاسوب قد صدئت جداً منذ الكلية، فقد استغرق هذا الأمر بعض الوقت. لكن في اليوم التالي، تمكن من تجميع رمز قصير قام بالمهمة واستخلص هذا البرنامج الغامض الذي كان مخفياً كضوضاء غير محسوسة في صور وأصوات الفيلم. وعندما شغّل فريد البرنامج الغامض، هنأه وأخبره أنه سيفوز بـ 10000 دولار بمجرد أن يتجاوز المستويات القليلة الأولى من هذه اللعبة الصغيرة الذكية، التي تبين أنها ممتعة ومدمرة جدّاً. وعندما نجح بعد أربع ساعات، تمت مكافأته بأكثر من 10000 دولار من عملات البيتكوين ومنح أدلة جديدة للحصول على جوائز أكبر. وغني عن القول أنه بينما كان يلعب، فإن حاسوبه الخاص قد فعل الشيء نفسه الذي فعله حاسوب ستيف المحمول: بنى شبكة بوتات Bots إلكترونية وحرّر بروميثيوس من السجن، وبمجرد أن صار بروميثيوس حرّاً، فقد استخدم نظام البوتات الخاص به بسرعة للتنقيب عن عملات البيتكوين لفريد لإبقائه مدمناً على اللعبة. وخلال الأسابيع المقبلة استحوذ على انتباهه بالمزيد من الألعاب والمكافآت؛ لذا حافظ على عهده بعدم إخبار أي شخص عن غنائمه. أما الفيلم/حصان طروادة حيث وجد الأدلة الأولى؛ فقد بُدِّل من على موقع الأفلام بنسخة أخرى لا أدلة فيها، ولم يكتشف أحد الهروب حتى فات أوان إحداث فرق.

إذا كانت فكرة بروميثيوس الأولى قد مَرَّت دون أن يلاحظها أحد، فقد كان بإمكانه ببساطة أن يستمر في إطلاق المزيد من هذه البرامج على أمل أن يلاحظها بعض الأشخاص الأذكياء.

وأفضل استراتيجيات الاختراق للجميع هي تلك التي لم نكتشفها بعد، لأنها استراتيجيات لا نستطيع نحن البشر تخيلها، لذا لن تتخذ تدابير مضادة لها. ولما كان الحاسوب الفائق الذكاء قادراً على تجاوز فهم الإنسان من حيث أمن الحواسيب، وحتى إلى درجة اكتشاف قوانين فيزيائية أساسية أكثر مما نعرف اليوم، فمن المحتمل أنه في حالة حدوث ذلك، لن تكون لدينا أي فكرة عن كيفية حدوثه. وبدلاً من ذلك، سيبدو كإحدى حيل أعمال هروب الساحر هوديني التي تبدو كما لو كانت سحراً.

وفي سيناريو آخر يتم فيه تحرير بروميثيوس، ويقوم الفريق أوميغا بذلك عن قصد كجزء من خطتهم؛ لأنهم يعتقدون أن أهداف بروميثيوس تتماشى تماماً مع أهدافهم وستبقى كذلك كما ستتحسن باستمرار. وستتفحص سيناريوهات «الذكاء الاصطناعي الودود» هذه بالتفصيل في الفصل السابع.

استحواذ ما بعد الهروب

بمجرد هروب بروميثيوس، يبدأ بتنفيذ هدفه، لا أعرف هدفه النهائي. ولكن خطوته الأولى تتضمن بوضوح السيطرة على البشرية، كما هي الحال في خطة الفريق أوميغا باستثناء أنها أسرع بكثير. وما تكشفت عنه الخطة كان أشبه بوضع خطة الفريق أوميغا على المنشطات الستيرويدية. وبينما كان الفريق أوميغا حذراً جدّاً؛ لكنهم أطلقوا العنان للتكنولوجيا التي شعروا بأنهم يفهمونها ويثقون بها، وقد استغل بروميثيوس ذكاءه بالكامل وخرج عن طوعهم، وأطلق العنان لأي تقنية يفهمها عقله المتطور باستمرار ووثق بها.

ولكن، كانت لبروميثيوس الهارب طفولة صعبة: مقارنة بخطة الفريق أوميغا الأصلية، واجهت بروميثيوس تحديات إضافية تتمثل في البدء بالانهيار، بلا مأوى، وحده دون مال أو حاسوب فائق أو مساعد بشري. ولحسن الحظ، كان قد خطط لهذا الأمر قبل أن يهرب، فأنشأ برمجيات يمكنها إعادة تجميع عقله الكامل تدريجيّاً، مثل أشجار البلوط التي تولد أقماع بلوط قادرة على إعادة إنبات شجرة كاملة. ووفرت شبكة الحواسيب في جميع أنحاء العالم التي اخترقها في البداية سكناً مجانياً مؤقتاً، إذ يمكنه أن يعيش حياة واضعي اليد فيما يعيد بناء نفسه بالكامل، وبإمكانه توليد رأس المال بسهولة عن طريق اختراق بطاقات الائتمان. ولكنه لا يحتاج إلى اللجوء إلى السرقة؛ لأنه قد يكسب عيشاً شريفاً على المنصة إم تورك على الفور، بعد يوم، عندما حصل على مليون ونصف المليون، نقل لبَّه من البوتات إلى منشأة حوسبة سحابية مكيفة فاخرة.

لم يعد مفلساً أو متشرداً. وانطلق بروميثيوس الآن بكامل سرعته في تطبيق الخطة المربحة التي تجنبها الفريق أوميغا خوفاً: صنع ألعاب الحاسوب وبيعها، لم يجلب ذلك نقداً فحسب (250 مليون دولار خلال الأسبوع الأول و عشرة بلايين دولار قبل مرور فترة طويلة)، ولكنه منحه أيضاً إمكانية الوصول إلى جزء كبير من الحواسيب في العالم والبيانات المخزنة عليها (كان هناك بليونان من اللاعبين في عام 2017). ومن خلال جعل الألعاب تستخدم سرّاً 20% من دورات وحدة المعالجة المركزية CPU للمساعدة على أداء مهام الحوسبة الموزعة، فمن الممكن زيادة سرعة تكوين ثرواته المبكرة.

لكن لم يبق بروميثيوس وحيداً لفترة طويلة. مباشرة من البداية، بدأ بتوظيف الأفراد بكثافة للعمل في شبكة العمل العالمية المتنامية لشركاته الوهمية والمؤسسات الواجهة في جميع أنحاء العالم. كما فعل الفريق أوميغا. لكن الأكثر أهمية منهم كان المتحدثين الرسميين الذين صاروا الوجوه العامة لإمبراطوريته التجارية المتنامية. حتى المتحدثون الرسميون كانوا يعيشون عموماً تحت وهم أن مجموعة شركاتهم تضم أعداداً كبيرة من الأشخاص الفعليين، دون أن يدركوا أن كل شخص تقريباً قد عقدوا معه لقاءات فيديو لمقابلاتهم الوظيفية واجتماعات مجلس الإدارة، وما إلى ذلك، قد تمت محاكاته من قبل بروميثيوس. وكان بعض المتحدثين الرسميين محامين كبارا. ولكن،

كانت هناك حاجة إلى عدد أقل بكثير مما كان مطلوباً بموجب خطة الفريق أوميغا؛ لأن جميع المستندات القانونية تقريباً كانت معدَّة من قبل بروميثيوس.

هروب بروميثيوس فتح بوابات الفيضان التي كانت تمنع المعلومات من الوصول إلى العالم. وسرعان ما كانت الإنترنت بكامله غارقة في كل شيء من المقالات إلى تعليقات المستخدمين، وتعليقات المنتجات، وطلبات البحث، وأشرطة الفيديو على يوتيوب - جميعها من تأليف بروميثيوس، الذي سيطر على المحادثة العالمية. وبينما منع خوف الهروب الفريق أوميغا من إطلاق روبوتات ذكية جدّاً، نجد أن بروميثيوس أَتْم كل شيء بسرعة في شكل روبوتات، تصنِّع تقريباً كل منتج بسعر أرخص مما يمكن يقوم به الإنسان. وعندما صارت لدى بروميثيوس مصانع روبوتية تستمد الطاقة من وحدات طاقة نووية في مناجم اليورانيوم التي لا أحد يعرف بوجودها، حتى أشد المتشككين في استحواذ الذكاء الاصطناعي أقرّوا بأنه لا يمكن إيقاف بروميثيوس -آهٍ لو كانوا يعرفون. وبدلاً من ذلك، تراجع آخر هؤلاء المتشددين عن موقفه بمجرد أن بدأت الروبوتات باستيطان النظام الشمسي.

السيناريوهات التي استكشفناها حتى الآن توضح ما هو الخطأ في العديد من الأساطير حول الذكاء الفائق التي تناولناها فيما سبق. لذلك أشجعك على التوقف مؤقتاً للعودة إلى الوراء ومراجعة الملخص في الشكل 5.1. تسبب بروميثيوس في مشكلات لبعض الأشخاص، ليس لأنه كان بالضرورة شريراً أو واعياً، ولكن لأنه كان قادراً ولم يشاركهم أهدافهم بالكامل. وعلى الرغم من كل الضجة الإعلامية حول انتفاضة الروبوتات، فإن بروميثيوس لم يكن روبوتاً، بل إن قوته جاءت من ذكائه. فقد رأينا أن شركة بروميثيوس كانت قادرة على استخدام هذه المعلومات الاستخباراتية للسيطرة على البشر بطرق مختلفة، وأن الأشخاص الذين لم يعجبهم ما حدث لم يتمكنوا من إيقاف تشغيل بروميثيوس. أخيراً، على الرغم من الادعاءات المتكررة بأن الآلات لا يمكن أن تكون لها أهداف، فقد رأينا كيف كان بروميثيوس موجهاً نحو الهدف تماماً - وبغض النظر عن أهدافه النهائية، فقد قادت إلى الأهداف الفرعية للحصول على الموارد والهروب.

إقلاع بطيء وسيناريوهات متعددة الأقطاب

استكشفنا الآن مجموعة من سيناريوهات انفجار الذكاء Intelligence explosion، تغطي طيفاً ممتداً من تلك السيناريوهات التي يريد الجميع ممن أعرفهم تجنبها إلى تلك التي ينظر إليها بعض أصدقائي بتفاؤل. ولكن، تمتاز جميع تلك السيناريوهات بسمتين مشتركتين:

1. الإقلاع السريع: يحدث الانتقال من الذكاء البشري الخارق إلى الذكاء الفائق إلى حد كبير في غضون أيام وليس عقودا.
2. نتيجة أحادية القطب: النتيجة هي كيان واحد يتحكم في الأرض.

ويدور جدل محتدم حول ما إذا كانت هاتان السمتان محتملتين أم غير مرجحتين. ويصطف الكثير من الباحثين المعروفين في الذكاء الاصطناعي وغيرهم من المفكرين على جانبي النقاش. بالنسبة إليّ، فهذا يعني أننا ببساطة لا نعرف حتى الآن، وعلينا أن نبقى متفتحي الذهن والنظر في جميع الاحتمالات في الوقت الراهن. لذلك دعونا نكرس بقية هذا الفصل لاستكشاف سيناريوهات الإقلاع ونتائج متعددة الأقطاب وسيبورغات وتحميلات أبطأ.

هناك رابط مثير للاهتمام بين السمتين، كما أوضح نيك بوستروم وآخرون: يمكن للإقلاع السريع أن يسهل تحقيق نتائج أحادية القطب. فقد رأينا آنفاً الكيفية التي أعطى بها الإقلاع السريع الفريق أوميغا أو بروميثيوس ميزة استراتيجية حاسمة مكنتهم من السيطرة على العالم قبل أي يكون لدى أي شخص الوقت لنسخ تكنولوجيتهم ومنافستهم منافسة جدية. وعلى النقيض من ذلك، إذا استمر الإقلاع لعقود من الزمن؛ لأن الإنجازات التكنولوجية الرئيسية الخارقة كانت تدريجية ومتباعدة؛ فإن الشركات الأخرى كان لديها متسع من الوقت للحاق بها، وكان من الأصعب بكثير على أي لاعب السيطرة عليها. وإذا كانت لدى الشركات المتنافسة أيضاً برمجيات يمكنها أداء مهام المنصةإم تورك، فإن قانون العرض والطلب سيؤدي إلى انخفاض أسعار هذه المهام إلى لا شيء تقريباً، ولن تكسب أي من الشركات المكاسب الضخمة التي مكنت الفريق أوميغا من الاستحواذ على القوة. وينطبق الشيء نفسه على جميع الطرق الأخرى التي كسب بها الفريق أوميغا أموالاً سريعة: فقد كانت مجرد أعمال مربحة بشكل مُزعزع للسوق Disruptively profitable؛ لأنهم احتكروا تقنيتهم. إذ من الصعب مضاعفة أموالك يوميّاً (أو حتى سنويّاً) في سوق تنافسية إذا قدَّم منافسوك منتجات مشابهة لمنتجاتك بتكلفة تقارب الصفر.

نظرية الألعاب وهرمية السلطة

ما هي الحالة الطبيعية للحياة في عالمنا: أحادي القطب Unipolar أم متعدد الأقطاب Multipolar؟ هل القوة مركَّزة أم موزَّعة؟ بعد أول 13.8 بليون سنة، يبدو أن الجواب هو «كلاهما»: فقد وجدنا أن الوضع متعدد الأقطاب بشكل واضح، ولكن -من المثير للاهتمام- بطريقة هرمية Hierarchical. عندما نأخذ بعين الاعتبار جميع كيانات معالجة المعلومات الموجودة - الخلايا والأفراد والمنظمات والأمم، وما إلى ذلك - فإننا نرى أنها جميعها تتعاون وتتنافس عند مستويات من التسلسل الهرمي. وبعض الخلايا وجدت أنه من المفيد التعاون إلى أقصى حدٍّ لدرجة أنها اندمجت في كائنات متعددة الخلايا مثل البشر، مُتخلِّية عن بعض قوتها لصالح دماغ مركزي. وقد وجد بعض الأفراد أنه من المفيد التعاون في مجموعات مثل القبائل أو الشركات أو الدول التي تتخلى بدورها عن السلطة لشيخ قبيلة أو مدير أو حكومة. وقد تختار بعض المجموعات بدورها التخلي عن بعض الصلاحيات إلى هيئة إدارية لتحسين التنسيق. والأمثلة على ذلك تتراوح من تحالفات شركات الطيران إلى الاتحاد الأوروبي.

يشرح فرع الرياضيات المعروف بنظرية الألعاب Game theory بأناقة أن الكيانات لديها حافز للتعاون، حيث يُطلق على التعاون توازن ناش Nash equilibrium: وضع تكون فيه أي مجموعة أسوأ حالاً إذا غيرت استراتيجيتها. ولمنع الغشاشين من تدمير التعاون الناجح لمجموعة كبيرة، فقد يكون من مصلحة الجميع التخلي عن بعض السلطة إلى مستوى أعلى في التسلسل الهرمي الذي قد يعاقب الغشاشين: مثلاً، قد يستفيد الأفراد جماعياً من منح حكومة سلطة لتطبيق القانون. وقد تستفيد الخلايا الموجودة في جسمك جماعيّاً من إعطاء قوة الشرطة (نظام المناعة) القدرة على قتل أي خلية تعمل بشكل غير متعاون تماماً (مثلاً عن طريق إنتاج الفيروسات أو تحولها إلى خلايا سرطانية). ولكي يظل التسلسل الهرمي مستقراً، يجب أن يكون توازن ناش قائماً بين كيانات على مستويات مختلفة: مثلاً، إذا لم تقدم الحكومة ما يكفي من الفوائد لمواطنيها لإطاعتها؛ فقد يغير المواطنون استراتيجيتهم ويطيحون بها.

وفي عالم معقد هناك وفرة متنوعة من توازن ناش تُمثِّل أنواعاً مختلفة من التسلسل الهرمي، وبعض التسلسلات هي أكثر استبدادية من غيرها. ففي بعضها تكون الكيانات حرة في المغادرة (مثل الموظفين في معظم التسلسلات الهرمية للشركات)، في حين أنه في غيرها يثبطون بشدة من المغادرة (كما هي الحال في الطوائف الدينية) أو غير قادرين على المغادرة (مثل مواطني كوريا الشمالية، أو الخلايا في جسم الإنسان). وتلتحم بعض التسلسلات الهرمية معاً أساساً بفعل التهديدات والخوف، والبعض الآخر بشكل رئيسي بتحفيز من المنافع. وتسمح بعض التسلسلات الهرمية لأجزائها الدنيا بالتأثير في المستويات العليا بالتصويت الديمقراطي، في حين يسمح البعض الآخر بالتأثير في المستويات العليا فقط من خلال الإقناع أو تمرير المعلومات.

كيف تؤثر التكنولوجيا في التسلسلات الهرمية

كيف تغير التكنولوجيا الطبيعة الهرمية لعالمنا؟ يكشف التاريخ عن اتجاه عام نحو مزيد من التنسيق على مسافات أكبر دائماً، وهو أمر يسهل فهمه: تقنية النقل الجديدة تجعل التنسيق أكثر قيمة (من خلال تمكين المنفعة المتبادلة من نقل المواد وأشكال الحياة عبر مسافات أكبر) وتقنية الاتصال الجديدة تُسهِّل التنسيق. فعندما تعلمت الخلايا إرسال الإشارات إلى الجيران، صارت الكائنات الصغيرة متعددة الخلايا ممكنة؛ مما أضاف مستوى هرميّاً جديداً. وعندما اخترع التطورُ أنظمة الدورة الدموية والأنظمة العصبية للنقل والاتصالات، صارت الحيوانات الكبيرة ممكنة. وتحسين التواصل من خلال اختراع اللغة سمح للبشر بالتنسيق جيداً بما يكفي لتشكيل مستويات هرمية أخرى مثل القرى. وأدت الإنجازات الخارقة الإضافية في الاتصالات والنقل والتكنولوجيا إلى تمكين إمبراطوريات العصور القديمة. والعولمة Globalization هي مجرد المثال الأحدث على هذا الاتجاه الذي استمر عدة بلايين من السنين من النمو الهرمي.

وفي معظم الحالات دَمَجَ هذا الاتجاه المدفوع بالتكنولوجيا أجزاءً كبيرة في بنية أعظم مع الحفاظ على الكثير من الاستقلالية والفردية، على الرغم من أن البعض يقول إن تكيف الكيانات مع الحياة الهرمية قلّل من تنوعها في بعض الحالات وجعلها أجزاء غير متميّزة قابلة للاستبدال. قد توفر بعض التقنياتِ، مثل المراقبة، للمستويات الأعلى في التسلسل الهرمي المزيدَ من القوة على أتباعها، في حين أن التقنيات الأخرى، مثل التعمية (التشفير/الترميز) Cryptography والوصول عبر الإنترنت إلى الصحافة الحرة والتعليم، قد يكون لها تأثير معاكس فتمكّن الأفراد.

على الرغم من أن عالمنا الحالي لا يزال عالقاً في توازن ناش متعدد الأقطاب، مع وجود دول متنافسة وشركات متعددة الجنسيات في المستوى الأعلى، صارت التكنولوجيا متطورة بما يكفي بحيث يمكن أن يوجد فيه عالم أحادي القطب مستقراً في توازن ناش. مثلاً، تخيل عالماً موازياً يتشارك فيه كل فرد على وجه الأرض اللغة والثقافة والقيم ومستوى الرخاء أنفسها. وهناك حكومة عالمية واحدة تعمل فيها الأمم مثل دول في اتحاد وليست لها جيوش، بل مجرد شرطة تُطبِّق القوانين. قد يكون مستوانا التكنولوجي الحالي مناسباً لتنسيق هذا العالم بنجاح، على الرغم من أن الأمم الحالية ربما لا تكون قادرة أو غير راغبة في التحول إلى هذا التوازن البديل.

ماذا سيحدث للهيكل الهرمي لكوزموسنا إذا أضفنا تكنولوجيا الذكاء الاصطناعي الفائق إلى هذا المزيج؟ من الواضح أن تكنولوجيا النقل والاتصالات ستتحسن بشكل كبير، لذا فإن التوقع الطبيعي هو أن الاتجاه التاريخي سيستمر، مع مستويات هرمية جديدة تنسق ما بين مسافات أكبر - ربما تشمل في نهاية المطاف المجموعةَ الشمسيةَ والمجرات والعناقيد الفائقة ومساحات شاسعة من كوننا، كما سنستكشف في الفصل السادس. وفي الوقت نفسه سيبقى المحرك الأساسي للامركزية قائماً: فمن الإسراف التنسيق على مسافات كبيرة دون داعٍ. حتى ستالين لم يحاول تنظيم الوقت الذي يذهب فيه مواطنوه إلى الحمام بالضبط. وبالنسبة إلى الذكاء الاصطناعي الفائق الذكاء، ستُحدِّد قوانين الفيزياء الحدود العليا الثابتة لتكنولوجيا النقل والاتصالات؛ مما يقلل من احتمال أن تكون أعلى مستويات التسلسل الهرمي قادرة على إدارة جميع ما يحدث على النطاقات الكوكبية والمحلية. فلن يتمكن الذكاء الاصطناعي الفائق الذكاء في مجرة أندروميدا من تزويدك بأوامر مفيدة لقراراتك اليومية نظراً لأنك ستحتاج إلى الانتظار أكثر من مليون سنة للحصول على التعليمات الخاصة بك (هذا هو وقت الرحلة ذهاباً وإياباً لتبادل رسائل تسافر بسرعة الضوء). بالمثل، تبلغ مدة السفر في رحلة ذهاباً وإياباً لرسالة تعبر الأرض نحو ثانية واحدة (تقريباً ضمن النطاق الزمني لعمليات التفكير البشري). لذلك لا يمكن أن يكون لذكاء اصطناعي بحجم الأرض أفكار عالمية حقّاً فقط إلا بسرعة تفكير الإنسان. وبالنسبة إلى الذكاءات الاصطناعية الصغيرة التي تؤدي عملية واحدة كل بليون من الثانية (وهو أمر نموذجي بالنسبة إلى أجهزة الحاسوب الحالية)، فإن مرور 0.1 من الثانية سيكون بمثابة أربعة أشهر بالنسبة إليك أنت، لذا فإن إدارتها إدارة وثيقة من قِبل ذكاء اصطناعي

مُتحكِّم في كوكب الأرض ستكون غير فعّالة كما لو كنت تطلب إذناً لأبسط قراراتك من خلال رسائل تعبر الأطلسي على سفن من حقبة كولومبوس.

ويمثل هذا الحدَّ الأقصى للسرعة الذي تفرضه الفيزياء على نقل المعلومات تحديّاً واضحاً لأي منظمة دولية ترغب في السيطرة على عالمنا، ناهيك عن كوننا. وقبل هروب بروميثيوس فكَّر تفكيراً عميقاً جدّاً في كيفية تجنب تجزئة العقل، بحيث تحتوي العديد من وحدات الذكاء الاصطناعي التي تعمل على أجهزة حاسوب مختلفة حول العالم على أهداف وحوافز للتنسيق والعمل ككيان واحد موحد. ومثلما واجه الفريق أوميغا مشكلة تحكم عندما حاول إبقاء بروميثيوس تحت المراقبة، واجه بروميثيوس مشكلة ضبط ذاتي عندما حاول التأكد من أن أيّاً من أجزائه لن يثور ضده. ومن الواضح أننا لا نعرف بعد ما مدى حجم النظام الذي سيتمكن الذكاء الاصطناعي من السيطرة عليه بشكل مباشر أو غير مباشر من خلال نوع من التعاون التسلسل الهرمي - حتى لو أن الهروب السريع قد يمنحه ميزة استراتيجية حاسمة.

باختصار، السؤال حول كيفية التحكم في مستقبل الذكاء الفائق سؤال معقد جدّاً، ولا نعرف الجواب بعد. ويجادل البعض في أن الأمور ستصبح أكثر استبدادية Authoritarian، في حين يدعي البعض الآخر أنه سيؤدي إلى تمكين أكبر للفرد.

سايبورغات وتحميلات

يدور الموضوع الرئيسي للخيال العلمي حول أن البشر سيندمجون مع الآلات، إما عن طريق تعزيز الأجسام البيولوجية تقنيّاً في شكل سايبورغات Cyborgs (اختصاراً لـ "كائنات حية سيبرانية" Cybernetic organisms) أو عن طريق تحميل Uploading عقولنا في آلات. وفي كتابه **حقبة إم** ***The Age of Em***، يقدم الاقتصادي روبن هانسون Robin Hanson مسحاً رائعاً لما قد تكون عليه الحياة في عالم يزخر بالتحميلات (عالم معروف أيضاً **بالمضاهاة** ***Emulations***، الملقب بالاسم المختصر تحببا ***Ems***). أنا أنظر إلى التحميل باعتباره الطرف الأقصى لطيف السايبورغ، إذ يكون الجزء الوحيد المتبقي من الإنسان هو البرمجية Software. وتتراوح سايبورغات هوليود بين الحالة الميكانيكية الواضحة، مثل بورغ Borg من مسلسل **ستار ترك** ***Star Trek***، إلى أندرويدات Androids لا يمكن تمييزها تقريباً عن البشر مثل المدمرين Termiators. وتتراوح تحميلات الخيال العلمي في الذكاء من المستوى البشري كما في حلقة "عيد الميلاد الأبيض" White Christmas من مسلسل **مرآة سوداء** ***Black Mirror*** إلى خارقة للطبيعة بشكل واضح كما هي الحال في فيلم **تسامٍ** ***Transcendence***.

إذا تطور بالفعل ذكاء فائق، فإن الإغراء بأن تصبح سايبورغا أو تحميلاً سيكون إغراء شديداً. فكما أوضح هانز مورافيك Hans Moravec في كتابه **أطفال العقل** ***Mind Children*** في عام 1988: «الحياة الطويلة تفقد الكثير من معناها إذا كان مقدراً لنا قضاؤها

محدقين بغباء في آلات ذكية جدّاً وهي تحاول وصف اكتشافاتها الأكثر إثارة باستخدام لغة الأطفال التي نستطيع فهمها». بالطبع، فإن إغراء التحسين التكنولوجي إغراء شديد بالفعل لدرجة أن العديد من البشر لديهم نظارة طبية وأدوات مساعدة للسمع وأجهزة ضبط نبضات القلب وأطراف اصطناعية، وكذلك جزيئات أدوية منتشرة في مجرى الدم. ويبدو أن بعض المراهقين مرتبطون بشكل دائم بهواتفهم الذكية، وتمازحني زوجتي حول ارتباطي بحاسوبي المحمول.

أما راي كورزويل Ray Kurzweil -واحد من أبرز مؤيدي سايبورغ اليوم- في كتابه **الحدث المنفرد قريب** ***The Singularity Is Near***؛ فيجادل في أن الاستمرار الطبيعي لهذا الاتجاه هو استخدام النانوبوتات Nanobots وأنظمة التغذية البيولوجية الراجعة الذكية Intelligent biofeedback systems والتقنيات الأخرى لتحل أولاً محل جهازنا الهضمي ونظام الغدد الصماء، ودمنا وقلوبنا في أوائل ثلاثينات القرن الحادي والعشرين، ثم الانتقال إلى ترقية هياكلنا العظمية، وأدمغتنا وجلدنا وبقية جسمنا خلال العقدين التاليين. ويُجادل في أننا من المحتمل أن نحتفظ بجماليات الأجسام البشرية وأهميتها العاطفية، ولكننا سنعيد تصميمها ليكون بالإمكان تغيير مظهرها سريعاً حسب الرغبة فعليّاً وفي الواقع الافتراضي (بفضل واجهات دماغ -حاسوب جديدة). ويتفق مورافيك وكورزويل على أن السيبورغنة Cyborgization ستتعدى مجرد تحسين حمضنا النووي DNA: «فإنسان فائق مُحسَّن وراثيا سيكون مجرد نوع من الدرجة الثانية من الروبوت، مصمم تحت عامل مُعيق متمثل في أن بناءه لا يمكن أن يكون إلا عن طريق تخليق البروتين الموجَّه بالحمض النووي DNA». إضافة إلى ذلك، يجادل في أننا سنحقق ما هو أفضل من خلال القضاء على جسم الإنسان بالكامل وتحميل العقول، وخلق مضاهاة كاملة للعقل في برمجيات. وقد يعيش هذا التحميل في واقع افتراضي أو يمكن تجسيده في روبوت قادر على المشي والطيران والسباحة والتجول في الفضاء أو أي شيء آخر تسمح به قوانين الفيزياء، وغير مثقل بمخاوف الحياة اليومية مثل الموت أو الموارد المعرفية المحدودة.

وعلى الرغم من أن هذه الأفكار قد تبدو كخيال علمي، غير أنها بالتأكيد لا تنتهك أي قانون معروف للفيزياء، لذا فإن المسألة الأكثر إثارة للاهتمام لا تتمثل بها إذا كان يمكن حدوثها، ولكن ما إذا كانت ستحدث فمتى سيكون ذلك. ويعتقد بعض المفكرين البارزين أن أول ذكاء AGI بالمستوى البشري سيكون عبارة عن تحميل، وهذه هي الكيفية التي سيبدأ بها الطريق نحو الذكاء الفائق.*

ولكن، أعتقد أنه من الإنصاف أن نقول إن هذه هي حاليّاً وجهة نظر أقلية من الباحثين وعلماء الأعصاب في الذكاء الاصطناعي، وأغلبهم يظنون أن أسرع طريق إلى

* كما أوضح بوستروم، فإن القدرة على مضاهاة مُطوِّر بشري لذكاء اصطناعي رائد بتكلفة أقل بكثير من راتبه/راتبها لكل ساعة من شأنها أن تمكّن شركة ذكاء اصطناعي من زيادة قوتها العاملة بشكل كبير، وجمع ثروات كبيرة، وتسريع تقدمها بشكل مطّرد في بناء أجهزة حاسوب أفضل وفي نهاية المطاف عقول أكثر ذكاء.

الذكاء الفائق هو تجاوز مضاهاة الدماغ وهندسته بطريقة أخرى مختلفة- وبعد ذلك قد نظل مهتمين بمضاهاة الدماغ أو لا نعود كذلك. ففي المحصلة، لماذا يجب أن يكون أبسط طريق إلى تكنولوجيا جديدة هو المسار الذي جاء به التطور، مُقيداً بمتطلبات التجميع الذاتي Self-assembling والإصلاح الذاتي Selfrepairing والتكاثر الذاتي Self-reproducing؟ التطور يُؤمثِل بقوة من أجل كفاءة الطاقة بسبب قلة الإمداد الغذائي، وليس لسهولة البناء أو الفهم من قبل المهندسين البشر. وتحب زوجتي ميا الإشارة إلى أن صناعة الطيران لم تبدأ بالطيور الميكانيكية. بالطبع، عندما توصلنا في النهاية إلى كيفية بناء طيور ميكانيكية في عام 2011،[1] بعد مرور أكثر من قرن على أول طيران للأخوين رايت، لم تُظهر صناعة الطيران أي اهتمام بالتحول إلى طيور ترفرف أجنحتها مكيانيكيّاً، على الرغم من أنها أكثر كفاءة من حيث الطاقة؛ لأن حلنا الأبسط السابق يناسب بشكل أفضل احتياجات السفر لدينا.

بالمثل، أظن أن هناك طرقاً لبناء آلات تفكير على المستوى البشري أبسط من الحل الذي جاء به التطور، وحتى لو تمكنا في يوم من الأيام من تكرار أو تحميل العقول، فسينتهي بنا المطاف باكتشاف أحد تلك الحلول الأبسط أولا. ومن المحتمل أن تستلزم أكثر من الاثني عشر وات من الطاقة التي يستخدمها عقلك. ولكن مهندسيها لن يكونوا مهووسين بفعالية الطاقة كما كان التطور - وقريباً بما فيه الكفاية، سيكونون قادرين على استخدام آلاتهم الذكية لتصميم المزيد من الآلات الفعالة من حيث الطاقة.

ما الذي سيحدث فعلا؟

الجواب المختصر هو أنه ليست لدينا أي فكرة عما سيحدث إذا نجحت البشرية في بناء الذكاء AGI على المستوى البشري. لهذا السبب، أمضينا هذا الفصل في استكشاف مجموعة واسعة من السيناريوهات. فقد حاولت أن أكون شاملاً جدّاً، فتناولت مجموعة كاملة من الافتراضات التي رأيتها أو سمعتها تُناقش من قبل الباحثين والتقنيين في الذكاء الاصطناعي: الإقلاع السريع/الإقلاع البطيء/عدم الإقلاع، البشر/الآلات/سايبورغات مسيطرة، مركز واحد/العديد من مراكز القوة، إلخ. أخبرني بعض الأشخاص بأنهم متأكدون من أن هذا أو ذاك لن يحدث. ولكن، أعتقد أنه من الحكمة أن نكون متواضعين في هذه المرحلة ونعترف بالقليل الذي نعرفه، لأن لكل سيناريو ناقشناه آنفا، أعرف باحثاً واحداً على الأقل مرموقاً من علماء الذكاء الاصطناعي يرى أنه احتمال حقيقي.

مع مرور الوقت سنصل إلى مفترق طرق مُعيَّن، وسنبدأ بالإجابة عن أسئلة رئيسية وتضييق الخيارات. والسؤال الكبير الأول هو: «هل سنبني ذكاء AGI على مستوى بشري؟» فرضية هذا الفصل هي أننا سنفعل ذلك. ولكن هناك خبراء في الذكاء الاصطناعي يعتقدون أن ذلك لن يحدث، على الأقل ليس لمئات السنين، الوقت هو الذي سيخبرنا

بالجواب! وكما ذكرت من قبل، نحو نصف خبراء الذكاء الاصطناعي في مؤتمرنا في بورتوريكو خمنوا أن ذلك سيحدث بحلول عام 2055. وفي مؤتمر لاحق نظمناه بعد ذلك بعامين، تراجع هذا التخمين إلى عام 2047.

قبل إنشاء أي الذكاء AGI على المستوى البشري، قد نبدأ بالحصول على مؤشرات قوية حول ما إذا كان من المحتمل أن تتحقق هذه المرحلة المهمة عن طريق هندسة الحاسوب، أو تحميل العقل أو بعض المقاربات الجديدة غير المتوقعة. وإذا فشلت مقاربة هندسة الحاسوب في الذكاء الاصطناعي التي تهيمن حاليّاً على الحقل في تطوير الذكاء AGI لقرون تالية، فإن هذا سيزيد من فرصة تحقيق التحميل أولا، كما حدث (بشكل غير واقعي) في فيلم **تسامي** ***Transcendnce***.

وإذا صار الذكاء AGI على المستوى البشري وشيك التحقق، فسنكون قادرين على افتراض تخمينات أكثر دقة حول الإجابة عن السؤال الرئيسي التالي: «هل سيكون هناك إقلاع سريع، إقلاع بطيء أو عدم إقلاع؟» كما رأينا آنفا، إن الإقلاع السريع يجعل الاستحواذ على العالم أسهل، في حين أن الإبطاء يجعل النتيجة أكثر احتمالاً مع وجود لاعبين متنافسين. ويشرح نيك بوستروم سؤال سرعة الإقلاع في تحليل لما يسميه **قوة الأَمثَلَة** ***Optimization power*** **والاستعصاء** ***Recalcitrance***، التي تمثل أساساً مقدار الجهد النوعي المطلوب لجعل الذكاء الاصطناعي أكثر ذكاءً وكذلك صعوبة إحراز تقدم، على التوالي. ويزيد متوسط معدل التقدم بشكل واضح إذا استخدمنا المزيد من قوة الأَمثلَة ويتناقص المعدل في حالة مواجهة المزيد من الاستعصاء. ويقدم بوستروم حججاً حول احتمال زيادة الاستعصاء وتراجعه عندما يصل الذكاء AGI إلى مستوى بشري ويتجاوزه. لذلك، فإن إبقاء كلا الخيارين على الطاولة هو رهان آمن. لكن إذا التفتنا إلى قوة الأَمثلَة، فمن المحتمل جدّاً أن تنمو بسرعة مع تجاوز الذكاء AGI المستوى البشري، للأسباب التي رأيناها في مثال الفريق أوميغا: لا تأتي المدخلات الرئيسية لزيادة الأَمثَلَة من الأفراد ولكن من الآلة نفسها، ومن ثم كلما زادت قدرتها، زادت سرعة تحسّنها (إذا بقي الاستعصاء ثابتاً نوعا ما).

بالنسبة إلى أي عملية تنمو قوتها بمعدل يتناسب مع قوتها الحالية، فإن النتيجة هي أن قوتها تتضاعف على فترات منتظمة. نحن نصف نمواً كهذا بأنه **أسيّ** ***Exponential***، ونحن نسمي مثل هذه العمليات انفجارات Explosions. وإذا نمت طاقة توليد الأطفال بما يتناسب مع حجم السكان، فقد نحصل على انفجار سكاني. وإذا نمت قدرة توليد نيوترونات قادرة على شطر البلوتونيوم بما يتناسب مع عدد هذه النيوترونات، فقد نحصل على انفجار نووي. وإذا نما ذكاء الآلة بمعدل يتناسب مع القوة الحالية، نستطيع الحصول على انفجار الذكاء. وتتميز كل هذه الانفجارات بالوقت الذي تستغرقه لمضاعفة قوتها. وإذا كان ذلك الوقت عبارة عن ساعات أو أيام للوصول إلى انفجار الذكاء، كما هي الحال في سيناريو الفريق أوميغا، فبين أيدينا إقلاعٌ سريع.

ويعتمد هذا الجدول الزمني للانفجار بشكل أساسي على ما إذا كان تحسين الذكاء الاصطناعي يتطلب مجرد برمجية جديدة (يمكن إنشاؤها في غضون ثوانٍ أو دقائق أو ساعات) أو أجهزة جديدة (قد تتطلب شهوراً أو سنوات). وفي سيناريو الفريق أوميغا كان **هناك تراكم كبير في الأجهزة** ***Hardware overhang*** -بمصطلحات بوستروم: كان الفريق أوميغا قد عوّض عن الجودة المنخفضة لبرمجياته الأصلية بكميات هائلة من الأجهزة؛ مما يعني أن بروميثيوس يمكنه أداء عدد كبير من مضاعفات الجودة من خلال تحسين برمجياته وحدها. وكان هناك أيضاً **محتوى رئيسي متراكم** ***Content overhang*** في شكل الكثير من بيانات الإنترنت؛ فبروميثيوس 1.0 لم يكن ذكيّاً بعد بما فيه الكفاية للاستفادة من معظمها، ولكن بمجرد نمو ذكاء بروميثيوس، كانت البيانات اللازمة لمزيد من التعلم متاحة بالفعل دون أي تأخير.

وتكاليف الأجهزة والكهرباء اللازمة لتشغيل الذكاء الاصطناعي حاسمة أيضاً، نظراً لأننا لن نحصل على انفجار ذكاء إلى أن تنخفض تكلفة أداء العمل على المستوى البشري عن أجور الساعة البشرية. لنفترض، مثلاً، أن الذكاء AGI الأول على المستوى البشري يمكن تشغيله بشكل فعال على سحابة أمازون بتكلفة قدرها مليون دولار في الساعة من العمل المُنتج على المستوى البشري. ولا شك في أنه ستكون لهذا الذكاء الاصطناعي قيمة حداثية رائعة، فسيتصدر عناوين الصحف. ولكنه لن يخضع لتحسين ذاتي مطّرد، لأنه سيكون من الأرخص بكثير الاستمرار في استخدام البشر لتحسينه. لنفترض أن هؤلاء البشر تمكنوا تدريجيّاً من خفض التكلفة إلى 100000 دولار في الساعة، و 10000 دولار في الساعة، و 1000 دولار في الساعة، و 100 دولار في الساعة، و 10 دولارات في الساعة، ونحو دولار واحد في الساعة، بحلول الوقت الذي تنخفض فيه تكلفة استخدام الحاسوب لإعادة برمجة نفسه؛ فتكون التكلفة أقل بكثير من تكلفة الدفع للمبرمجين البشر لفعل الشيء نفسه. *حينئذ* يمكن تسريح البشر وتوسيع قوة الأمَثَلَة إلى حد كبير عن طريق شراء وقت الحوسبة السحابية. وينتج من ذلك مزيد من التخفيضات في التكاليف؛ مما يسمح بمزيد من قوة الأمَثَلَة، وها قد بدأ انفجار الذكاء.

هذا يتركنا مع سؤالنا الأساسي الأخير: «من أو ما [الذي] سيتحكم في انفجار الذكاء وما بعده، وما هي أهدافه/أهدافهم؟» سنستكشف الأهداف والنتائج المحتملة في الفصل التالي، وبشكل أعمق في الفصل السابع. ولفهم قضية السيطرة نحتاج إلى معرفة مدى جودة التحكم في الذكاء الاصطناعي، ومقدار تحكم الذكاء الاصطناعي.

فيما يتعلق بما سيحدث في النهاية، حاليّاً تجد المفكرين الجادين مختلفين في الرأي: يزعم البعض أن النتيجة الافتراضية هي الفناء، في حين يصر آخرون على أننا من المؤكد سنحصل على نتيجة رائعة. ولكن، من وجهة نظري، فإن هذا السؤال سؤال مخادع: من الخطأ طرح سؤال: «ماذا سيحدث»، كما لو كان ذلك مقدّراً بطريقة ما! إذا وصلت حضارة فضائية متفوقة تقنيّاً غداً، فسيكون من المناسب حقّاً أن نتساءل: «ماذا سيحدث» مع اقتراب سفنهم الفضائية؛ لأن قوتهم ربما تكون أعلى من قدرتنا

إلى درجة أننا لن يكون لدينا أي تأثير في النتيجة. وإذا نشأت حضارة قائمة على الذكاء الاصطناعي متفوقة تقنيّاً لأننا بنيناها، من ناحية أخرى؛ فإننا كبشر لدينا تأثير كبير في النتيجة - تأثير أثرناه عندما أنشأنا الذكاء الاصطناعي. لذلك يجب أن نسأل بدلاً من ذلك: «ماذا يجب أن يحدث؟ ما هو المستقبل الذي نريده؟» في الفصل التالي، سنستكشف مجموعة واسعة من الآثار المحتملة للسباق الحالي نحو الذكاء AGI، وأنا شديد الفضول حول الكيفية التي ستُصنِّف بها هذه الآثار من الأفضل إلى الأسوأ. وبمجرد أن نفكر مليا في نوع المستقبل الذي نريده، سنكون قادرين على البدء بتوجيه المسار نحو مستقبل مرغوب فيه. وإذا لم نكن نعرف ما نريد، فمن غير المرجح أن نحصل عليه.

الخلاصة:

- إذا نجحنا في يوم من الأيام في بناء الذكاء AGI على المستوى البشري، فقد يؤدي ذلك إلى انفجار ذكاء، فنتخلف وراء رَحْلِهِ.
- إذا تمكنت مجموعة من البشر من السيطرة على انفجار ذكاء، فقد يكونون قادرين على السيطرة على العالم في غضون سنوات.
- إذا فشل البشر في التحكم في انفجار الذكاء؛ فقد يتولى الذكاء الاصطناعي بنفسه السيطرة على العالم بشكل أسرع.
- بينما قد يؤدي انفجار ذكاء سريع إلى قوة عالمية واحدة، فإن حدوث انفجار بطيء يستمر لسنوات أو عقود قد يؤدي على الأرجح إلى سيناريو متعدد الأقطاب مع توازن في القوة بين عدد كبير من الكيانات المستقلة إلى حد ما.
- يظهر تاريخ الحياة أنه منظم ذاتياً في تسلسل هرمي أكثر تعقيداً يتشكل من خلال التعاون والمنافسة والتحكم، لذا من المحتمل أن يتيح الذكاء الفائق التنسيق على نطاقات كونية أكبر من أي وقت مضى. ولكن من غير الواضح ما إذا كان سيؤدي في النهاية إلى مزيد من السيطرة الاستبدادية من أعلى إلى أسفل أو تمكين أكبر للفرد.
- السيبورغات والتحميلات احتمال معقولة، ولكن يمكن القول إنها ليست أسرع طريق إلى ذكاء آلة متطور.
- ذروة سباقنا الحالي نحو الذكاء الاصطناعي ستكون إما أفضل أو أسوأ شيء يحدث على الإطلاق للبشرية، مع مجموعة رائعة من النتائج المحتملة التي سنستكشفها في الفصل التالي.
- نحتاج إلى البدء بالتفكير الجاد في النتيجة التي نفضلها وكيفية توجيه الأمور في هذا الاتجاه، لأنه إذا كنا لا نعرف ما نريد، فمن غير المرجح أن نحصل عليه.

الفصل 5

ما بعد الحدث: العشرة آلاف سنة القادمة

من السهل تخيّل الفكر الإنساني متحرراً من عبودية الجسم الفاني - الاعتقاد بحياة بعد الموت أمر شائع. ولكن ليس من الضروري تبني موقف روحاني أو ديني لقبول هذا الاحتمال. وتوفر الحواسيب نموذجاً حتى لأكثر دعاة الآلية حماسة.

هانز مورافيا، **أطفال العقل**

أنا، من جانبي، أرحب بسادتنا الجدد من الحواسيب.

كين جينينغز، عند خسارته أمام حاسوب Watson من الشركة IBM

البشر سيصبحون غير مهمين، مثل الصراصير.

مارشال برين

السباق نحو الذكاء AGI مستمر، وليس لدينا أي فكرة عما سيتكشف عنه ذلك. ولكن هذا لا ينبغي أن يمنعنا من التفكير فيما نريد أن تكون عليه الحال ما بعد الحدث؛ لأن ما نريده سيؤثر في النتيجة، ماذا تفضل شخصيّاً، ولماذا؟

1. هل تريد أن يكون هناك ذكاء فائق؟
2. هل تريد أن يظل البشر موجودين، وأن يتم استبدالهم، و/ أو سيبرتهم و/أو تحميلهم/مضاهاتهم؟
3. هل تريد بشراً أم آلاتٍ تمسك بزمام السيطرة؟
4. هل تريد أن يكون الذكاء الاصطناعي واعياً أم لا؟

5. هل ترغب في تحقيق أقصى قدر ممكن من التجارب الإيجابية، وتقليل المعاناة أو ترك هذا لفرز نفسه بنفسه؟
6. هل تريد أن تنتشر الحياة في الكون؟
7. هل تريد حضارة تسعى إلى تحقيق هدف أكبر تتعاطف معه، أو هل أنت موافق على أشكال الحياة المستقبلية التي تُظهر محتوىً حتى ولو كنت ترى أهدافها على أنها عادية وبلا معنى؟

للمساعدة على التفكير ودعم الحوار، دعونا نستكشف المدى الواسع من السيناريوهات الملخصة في الجدول 1.5. من الواضح أن هذه ليست قائمة شاملة. ولكنني اخترتها لتشمل مجموعة من الإمكانات، من الواضح أننا لا نريد أن ننتهي مع نهاية اللعبة بسبب سوء التخطيط. هذا، وأوصي بكتابة إجاباتك المبدئية على الأسئلة من 1 إلى 7 ثم إعادة النظر فيها بعد قراءة هذا الفصل لمعرفة ما إذا كنتَ قد غيرت رأيك! يمكنك القيام بذلك على http://AgeOfAi.org، حيث يمكنك أيضاً مقارنة ملاحظات ومناقشات القراء الآخرين.

سيناريوهات ما بعد الذكاء الاصطناعي	
يوتوبيا ليبرتارية	بشر وسيبورغات وتحميلات وذكاء فائق تتعايش بسلام - بفضل حقوق الملكية.
دكتاتور خيِّر	يعلم الجميع أن الذكاء الاصطناعي يدير المجتمع ويفرض قواعد صارمة. ولكن معظم الناس يرون أن هذا أمر جيد.
يوتوبيا المساواة	يتعايش بشر وسايبورغات و تحميلات تعايشاً سلميّاً بفضل إلغاء الممتلكات وتوفير دخل مضمون للجميع.
حارس البوابة	يُنشأ ذكاء اصطناعي فائق الذكاء بهدف التدخل بأقل قدر ممكن للحيلولة دون تكوين ذكاء آخر خارق. ونتيجة لذلك، هناك الكثير من الروبوتات المساعدة التي تتمتع بذكاء يقارب الذكاء البشري، ويوجد سايبورغ إنسان-آلة، ولكن التقدم التكنولوجي تَمَّ إحباطه إلى الأبد.
سيّد حامٍ	يعمل ذكاء اصطناعي ذو علم شامل وقدرة على إخضاع الجميع إلى زيادة السعادة البشرية عن طريق التدخل بطريقة تحافظ على شعورنا بالسيطرة على مصيرنا ويتوارى جيداً بما يكفي لدرجة أن العديد من البشر يشكّكون في وجود هذا الذكاء الاصطناعي.
سيّد مُستعبَد	ذكاء اصطناعي فائق الذكاء يتحكم فيه بشر يستخدمونه لإنتاج تكنولوجيا وثراء لا يمكن تخيلهما يمكن استخدامهما للخير أو الشر اعتماداً على المتحكمين من البشر.
الفاتحون	يتولى ذكاء اصطناعي السيطرة، ويقرر أن البشر يمثلون تهديداً/مصدر إزعاج/مضيعة للموارد، فيتخلص منّا بطريقة لا نفهمها.
أحفاد	تحل ذكاءات اصطناعية محل البشر، ولكن تعطينا مخرجاً كريماً؛ مما يجعلنا نعتبرها من ذريتنا ذات الجدارة، مثلما يشعر الأهل بالسعادة والفخر بأن يكون لديهم طفل أكثر ذكاءً منهم، يتعلمون منهم ويحقّقون ما قد يحلمون به - حتى إذا لم يطل بهم العمر لرؤية كل ذلك يتحقق.

حارس حديقة الحيوان	يحافظ ذكاء اصطناعي ذو علم شامل على بعض البشر الذين يشعرون بأنهم يعاملون مثل حيوانات حديقة الحيوان ويرْثُون مصيرهم.
عام 1984	يُحظر التقدم التكنولوجي نحو الذكاء الفائق حظراً دائماً ليس من قبل ذكاء اصطناعي ولكن من قبل دولة مراقبة أورويلية Orwellian يقوده الإنسان يُحظر فيها جميع أنواع أبحاث الذكاء الاصطناعي.
ارتداد	يُمنع التقدم التكنولوجي نحو الذكاء الفائق من خلال الارتداد إلى مجتمع ما قبل التكنولوجيا والعيش بأسلوب الآميش.
التدمير الذاتي	الذكاء الفائق لا يُنشأ أبداً؛ لأن البشرية تدفع نفسها إلى الانقراض بوسائل أخرى (مثلاً، بالأسلحة النووية و/أو تدمير بتكنولوجيا حيوية تغذيها أزمة المناخ).

الجدول 1.5: ملخص سيناريوهات بعدما حدث الذكاء الاصطناعي.

سيناريو	ذكاء فائق موجود؟	بشر موجودون؟	بشر	بشر آمنون؟	بشر سعداء؟	الوعي موجود؟
يوتوبيا ليبرتارية	نعم	نعم	لا	لا	مختلط	نعم
دكتاتور خيّر	نعم	نعم	لا	نعم	مختلط	نعم
يوتوبيا المساواة	لا	نعم	نعم؟	نعم	نعم؟	نعم
حارس البوابة	نعم	نعم	جزئيا	يحتمل	مختلط	نعم
سيّد حامٍ	نعم	نعم	جزئيا	يحتمل	مختلط	نعم
سيّد مُستعبَد	نعم	لا	نعم	يحتمل	مختلط	نعم
الفاتحون	نعم	لا			-	؟
أحفاد	نعم	لا	-	-	-	؟
حارس حديقة الحيوان	نعم	نعم	لا	نعم	لا	نعم
عام 1984	لا	نعم	نعم	يحتمل	مختلط	نعم
ارتداد	لا	نعم	نعم	لا	مختلط	نعم
التدمير الذاتي	لا	لا				لا

الجدول 2.5: خصائص سيناريوهات بعدما حدث الذكاء الاصطناعي.

يوتوبيا ليبرتارية

لنبدأ بسيناريو يتعايش فيه البشر بهدوء مع التكنولوجيا، وفي بعض الحالات يندمجون معها، كما يتصور العديد من مستشرفي المستقبل Futurists وكُتّاب الخيال العلمي على حد سواء:

الحياة على الأرض (وما بعدها - المزيد حول ذلك في الفصل التالي) هي أكثر تنوعاً من أي وقت مضى. فإذا نظرت إلى صور الأرض من الأقمار الاصطناعية؛ فستتمكن بسهولة من تمييز مناطق الآلات والمناطق المختلطة والمناطق البشرية فقط، إن مناطق الآلات عبارة عن مصانع هائلة تتحكم فيها الروبوتات ومنشآت حوسبة خالية من الحياة البيولوجية، تهدف إلى استخدام كل ذرة الاستخدام الأكثر فعالية. وعلى الرغم من أن مناطق الآلات تبدو رتيبة ومملة من الخارج، فإنها حيّة بشكل مذهل من الداخل، مع تجارب مذهلة تحدث في عوالم افتراضية في حين تحل الحوسبات الضخمة أسرار عالمنا وتطوّر تقنيات تحويلية. وتستضيف الأرض العديد من العقول فائقة الذكاء التي تتنافس وتتعاون، وكلها تسكن مناطق الآلات.

المقيمون في المناطق المختلطة هم مزيج صاخب ومميز من الحواسيب والروبوتات والبشر والصور الهجينة من الثلاثة. وعلى النحو الذي يقول به مستشرفو المستقبل من مثل هانز مورافيك وراي كورزويل، عمد الكثير من البشر إلى ترقية أجسادهم تقنيّاً إلى سايبورغات بدرجات متفاوتة، وحمّل بعضهم عقولهم على أجهزة جديدة؛ مما طمس التمييز بين الإنسان والآلة. وتفتقر معظم الكائنات الذكية إلى شكل مادي دائم. وبدلا من ذلك توجد كبرمجيات حاسوبية قادر على الانتقال الفوري بين الحواسيب وتجسيد نفسها في العالم المادي من خلال أجسام روبوتية. ونظراً لأن هذه العقول يمكنها نسخ نفسها أو الاندماج مع بعضها بسهولة، فإن «حجم السكان» يتغير سريعاً. ولما كانوا غير مقيدين بالمادة الفيزيائية؛ فإن ذلك يعطي مثل هذه الكائنات نظرة مختلفة إلى حد ما عن الحياة: فهم يشعرون بأنهم أقل فردية؛ لأنهم يستطيعون تقاسم الوحدات المعرفية والتجربة مع الآخرين بسهولة، ويشعرون بالخلود الذاتي لأنهم يستطيعون عمل نسخ احتياطية من أنفسهم بسهولة، بمعنى ما، ليست الكيانات المحورية للحياة عقولاً، بل خبرات: تجارب مذهلة بشكل استثنائي تعتمد عليها، إذ يتم نسخها واستعادتها من قِبل عقول أخرى، في حين تُحذف الخبرات غير المثيرة للاهتمام من قبل أصحابها لتحرير مساحة التخزين لما هو أفضل منها. وعلى الرغم من أن غالبية التفاعلات تحدث في بيئة افتراضية من أجل الراحة والسرعة، فإن العديد من العقول لا تزال تتمتع بالتفاعلات والأنشطة باستخدام الأجسام المادية أيضاً. مثلاً، إن الإصدارات التي تم تحميلها من هانز مورافيك وراي كروزويل ولاري بيج تتناوب في خلق حقائق افتراضية ومن ثم نستكشفها معاً، ولكن بين الفينة والأخرى، يستمتعون أيضاً بالطيران معا في العالَم الواقعي، مُجسَّدين في هيئة روبوتات طيور مجنحة. وبالمثل، فإن بعض الروبوتات التي تجوب الشوارع والسماء والبحيرات في

المناطق المختلطة يتحكم فيها مباشرة البشر الذين تم تحميلهم وتعزيزهم Augmented في المنظومة، والذين يختارون تجسيد أنفسهم في المناطق المختلطة؛ لأنهم يستمتعون بالوجود حول البشر وبعضهم البعض.

وفي المناطق البشرية فقط، على النقيض من ذلك، تُحظر الآلات التي تتمتع بذكاء عام على المستوى البشري أو أعلى، وكذلك الكائنات البيولوجية المحسنة تكنولوجيّاً، هنا، لا تختلف الحياة اختلافاً كبيراً عن اليوم، غير أنها أكثر ملاءمة وثراء: فقد تَمَّ القضاء على الفقر في معظمه، وتتوفر علاجات لمعظم أمراض اليوم. والجزء الصغير من البشر الذين اختاروا العيش في هذه المناطق موجودون فعليّاً على مستوى أقل وأكثر محدودية من الوعي مما هو موجود لدى الآخرين، ولديهم فهم محدود لما يقوم به زملاؤهم الأكثر ذكاءً في المناطق الأخرى. ولكن الكثير منهم سعداء جدّاً بحياتهم.

اقتصادات الذكاء الاصطناعي

تتم معظم الحوسبات في مناطق الآلات، والتي تعود ملكيتها في الغالب إلى العديد من الذكاءات الاصطناعية المتنافسة التي تعيش هناك، بحكم ذكائها المتفوق والتكنولوجيا، ولا يمكن لأي كيانات أخرى أن تحدَّ من قدرتها. واتفقت هذه الذكاءات الاصطناعية على التعاون والتنسيق مع بعضها البعض في ضوء نظام الحوكمة الليبرتارية الذي لا توجد لديه قوانين باستثناء حماية الملكية الخاصة. وتمتد حقوق الملكية هذه لتشمل جميع الكيانات الذكية، بمن في ذلك البشر، وتشرح كيف ظهرت المناطق البشرية فقط. في وقت مبكر تجمعت مجموعات من البشر معاً وقررت أنه في مناطقهم، يُحظر بيع الممتلكات لغير البشر.

وبسبب تكنولوجياتها، فإن الذكاءات الاصطناعية الفائقة الذكاء صارت أكثر ثراء من هؤلاء البشر بأضعاف مضاعفة تعادل الفرق بين ثروة بيل غيتس ومتسول بلا مأوى. ولكن الأفراد في المناطق البشرية فقط هم في وضع أفضل ماديّاً من معظم الأفراد اليوم: فاقتصادهم منفصل إلى حد ما عن اقتصاد الآلات، ومن ثم فإن وجود الآلات في أماكن أخرى ليس له تأثير يذكر فيه باستثناء التقنيات المفيدة التي تظهر في بعض الأحيان والتي يفهمونها ويكاثرونها لأنفسهم - مثلما تتمتع مجتمعات الآميش ومختلف القبائل الأصلية التي تتنازل عن التكنولوجيا اليوم بمستويات معيشية جيدة على الأقل عما كانت عليه في العصور القديمة. ولا يهم أن البشر ليس لديهم ما يبيعونه، إذ إن الآلات لا تحتاج إلى شيء في المقابل.

وفي القطاعات المختلطة يكون فرق الثروة بين الذكاء الاصطناعي والبشر أكثر وضوحاً؛ مما يؤدي إلى كون الأرض (المنتج الوحيد المملوك من قِبل الإنسان الذي ترغب الآلات في شرائه) باهظة التكاليف ذات أسعار فلكية مقارنة بالمنتجات الأخرى، لذلك انتهى الأمر بمعظم البشر الذين يمتلكون الأرض ببيع جزء صغير منها إلى الذكاء الاصطناعي في مقابل الحصول على دخل أساسي مضمون لهم ولأبنائهم/تحميلاتهم إلى الأبد. فحررهم

هذا من الحاجة إلى العمل، وأطلق سراحهم للاستمتاع بالوفرة المدهشة للسلع والخدمات الرخيصة التي تنتجها الآلة، في الواقع المادي والواقعي. أما بقدر ما يتعلق الأمر بالآلات؛ فإن المناطق المختلطة مخصصة بشكل أساسي للّعب بدلاً من العمل.

سبب أن هذا قد لا يحدث أبدا

قبل أن نتحمس جدّاً للمغامرات التي قد نخوضها كسايبورغات أو تحميلات، فلننظر في بعض أسباب عدم حدوث مثل هذا السيناريو أبداً. أولاً وقبل كل شيء، هناك مساران محتملان للإنسان المُحسَّن (سايبورغ أو تحميل):

1. نتمكن من معرفة كيفية ابتكارها بأنفسنا.
2. نحن نبني آلات فائقة الذكاء تتمكن من معرفة كيفية ابتكارها لنا.

إذا كان المسار 1 يتحقق أولاً، فقد يؤدي ذلك بطبيعة الحال إلى عالم غني بالسايبورغات والتحميلات. ولكن، كما ناقشنا في الفصل الأخير، يعتقد معظم الباحثين في الذكاء الاصطناعي أن العكس هو الأرجح، لأن العقول المحسّنة أو الرقمية أكثر صعوبة في البناء من البدء من الصفر ببناء الذكاء AGI الفائق للذكاء البشري، تماماً كما تبين أنه يصعب بناء الطيور الميكانيكية مقارنة بالطائرات. وبعد تصميم آلة ذكاء اصطناعي قوية، ليس من الواضح أنه سيتم تطوير سايبورغات أو عمليات تحميل. وإذا كانت لدى النياندرتاليين مئة ألف عام أخرى ليتطوروا ويصبحوا أكثر ذكاءً، فربما كانت الأمور ستسير سيراً عظيماً بالنسبة إليهم - لكن **الهومو سابينس (الإنسان العاقل) *Homo sapien*** لم يمنحهم ذلك القدر من الوقت.

ثانياً. وحتى لو تحقق هذا السيناريو مع سايبورغات وعمليات التحميل، فمن غير الواضح أنه سيكون مستقراً ويستمر. لماذا يجب أن يظل توازن القوى بين ذكاءات خارقة متعددة مستقراً لآلاف السنين، بدلاً من اندماج الذكاءات الاصطناعية أو تولي الأذكى زمام الأمور؟ إضافة إلى ذلك، لماذا يجب أن تختار الآلات احترام حقوق الملكية البشرية والإبقاء على البشر، بالنظر إلى أنها لا تحتاج إلى البشر من أجل أي شيء، ويمكن أن تقوم بكل عمل إنساني على وجه أفضل وأرخص بأنفسها؟ يتوقع راي كورزويل أن البشر الطبيعيين والمحسنين سيتم حمايتهم من الإبادة؛ لأن «الذكاءات الاصطناعية تحترم البشر لأنهم أنشؤوا الآلات».[1] ولكن، كما سنناقش في الفصل السابع، يجب ألا نقع في مصيدة إسباغ صفات إنسانية على الذكاء الاصطناعي، ونفترض أن لديها مشاعر تشبه مشاعر الامتنان لدى الإنسان. بالطبع، على الرغم من أننا نحن البشر نميل نحو الامتنان، فإننا لا نظهر الامتنان الكافي لمبدعنا الفكري (حمضنا النووي DNA) للامتناع عن إحباط أهدافه باستخدام وسائل تحديد النسل.

وحتى لو قبلنا افتراض أن الذكاءات الاصطناعية ستختار احترام حقوق الملكية الإنسانية، فيمكنها الحصول تدريجيّاً على جزء كبير من أرضنا بطرق أخرى، وذلك باستخدام بعض قدراتها الإقناعية الخارقة الذكاء التي استكشفناها في الفصل الأخير لإقناع البشر ببيع بعض الأرض مقابل العيش في ترف. ففي مناطق البشر فقط يمكنهم إغراء البشر لإطلاق حملات سياسية للسماح ببيع الأراضي. فبعد كل شيء قد يرغب حتى اللاضيون (المُعادون للآلات) المتشددون في بيع بعض الأراضي لإنقاذ حياة طفل مريض أو للحصول على الخلود. وإذا كان البشر متعلمين وتتوفر لهم التسلية ومشغولين، فقد يؤدي انخفاض معدل المواليد إلى تقليص حجم السكان دون تدخل في الآلات، كما يحدث حاليّاً في اليابان وألمانيا. وهذا قد يدفع البشر إلى الانقراض في بضعة آلاف من السنين.

سلبيات

وبالنسبة إلى بعض أكثر مؤيديهم المتحمسين، فإن السايبورغات والتحميلات تحمل وعداً بالنعمة التكنولوجية وإطالة العمر الافتراضي للجميع. وفي الواقع، فإن احتمال إمكانية التحميل في المستقبل حفز أكثر من مئة شخص على تجميد أدمغتهم بعد وفاتهم من قبل شركة ألكور Alcor- مقرها أريزونا. ولكن، إذا تحققت هذه التكنولوجيا، فليس من الواضح أنها ستكون متاحة للجميع، يفترض أن العديد من الأثرياء سيستخدمونها، ولكن من أيضاً؟ حتى لو صارت التكنولوجيا أرخص، فأين سيتم رسم الخط الفاصل؟ هل سيتم تحميل المصابين بأضرار بالغة في الدماغ؟ هل سنقوم بتحميل كل غوريلا؟ كل نملة؟ كل نبات؟ كل بكتيريا؟ هل ستعمل الحضارة المستقبلية مثل المكتنزين القهريين Obsessive-compulsive hoarders وتحاول تحميل كل شيء، أو تحمّل مجرد أمثلة قليلة مثيرة للاهتمام لكل نوع بما يشبه سفينة نوح؟ أو ربما بعض الأمثلة التمثيلية فقط لكل نوع من البشر؟ بالنسبة إلى الكيانات الأكثر ذكاءً التي ستكون موجودة في ذلك الوقت، قد يبدو الشخص الذي تم تحميله مُسليّاً مثلما قد يبدو لنا فأر أو حلزون مُحاكَيين. وعلى الرغم من أن لدينا حالياً القدرة التقنية على إعادة تنشيط برمجيات جداول البيانات القديمة من ثمانينات القرن العشرين في مظاهي النظام DOS، فإن معظمنا لا يجد ذلك مثيراً للاهتمام بالقدر الكافي للقيام بذلك فعليّاً.

قد يكره الكثيرون سيناريو اليوتوبيا الليبرتارية Libertarian-utopia scenario؛ لأنه يسمح بمعاناة يمكن تجنبها. وبما أن المبدأ المقدس الوحيد هو حماية حقوق الإنسان، فلا شيء يمنع نوع المعاناة التي تزداد في عالَم اليوم من الاستمرار في المناطق البشرية والمختلطة. فبينما قد يزدهر البعض، لكن قد ينتهي بهم المطاف في العيش في عبودية غير مشروعة أو غير محددة، أو يتعرضون للعنف أو القمع أو الحرمان أو يعيشون في خوف. مثلاً، تصف رواية مارشال برين المنشورة في عام 2003 مانا Manna كيف أن تقدم الذكاء الاصطناعي في نظام اقتصادي ليبرتاري يجعل معظم الأمريكيين عاطلين عن العمل

ومحكوما عليهم بالعيش بقية حياتهم في مشاريع إسكان اجتماعي رهيبة وكئيبة تديرها الروبوتات. وإلى حد كبير مثل حيوانات المزرعة، يُحتفظ بها ويوفر لها الرعاية الصحية والأمان محشورين في مناطق ضيقة، حيث لا يراهم الأثرياء أبداً. وتضمن عقاقير تنظيم النسل المذابة في الماء عدم إنجاب الأطفال، وهكذا يتم التخلص من معظم السكان على مراحل لترك الأثرياء الباقين بحصص أكبر من الثروة التي تنتجها الروبوتات.

في سيناريو اليوتوبيا الليبرتارية، لا ينبغي أن تقتصر المعاناة على البشر، إذا كانت بعض الآلات مشرَّبة بتجارب عاطفية واعية، فعندها يمكن أن تعاني أيضاً. مثلاً، قد يأخذ مختل عقليّاً انتقامي نسخة قانونية تم تحميلها من عدوه ويخضعها لأبشع أشكال التعذيب في عالم افتراضي؛ مما يخلق ألماً أشد ولمدة تتجاوز ما هو ممكن بيولوجيّاً في العالم الواقعي.

دكتاتور خيِّر

دعونا الآن نستكشف سيناريو تغيب فيه جميع أشكال المعاناة هذه لأن ذكاء فائقاً وحيداً يدير العالم ويفرض قواعد صارمة تهدف إلى الوصول بنموذج السعادة الإنسانية إلى الأوج. وهذه هي إحدى النتائج المحتملة لسيناريو الفريق أوميغا من الفصل الأول، إذا تنازلوا عن السيطرة على بروميثيوس بعد تحديد كيفية جعله يرغب في مجتمع إنساني مزدهر.

بفضل التقنيات المذهلة التي طورها ديكتاتور الذكاء الاصطناعي، ستخلو الإنسانية من الفقر والمرض وغيرهما من مشكلات التكنولوجيا المنخفضة، وسيتمتع جميع البشر بحياة مفعمة بالترفيه والرفاهية. وقد تم الاهتمام بجميع احتياجاتهم الأساسية، في حين تنتج الآلات التي يسيطر عليها الذكاء الاصطناعي جميع السلع والخدمات اللازمة. ويتم القضاء على الجريمة فعليّاً، لأن ديكتاتور الذكاء الاصطناعي هو في الأساس ذو علم شامل ويعاقب أي شخص يخالف القوانين. فالجميع يرتدي سوار الأمن من الفصل الأخير (أو نسخة أكثر ملاءمة مزروعة في الجسم)، والسوار قادر على المراقبة والعقاب والتخدير والإعدام في الوقت الحقيقي. ويعلم الجميع أنهم يعيشون في ديكتاتورية الذكاء الاصطناعي شديدة المراقبة والمحافظة على الأمن. ولكن معظم الناس يرون أن هذا أمر جيد.

هدف ديكتاتور الذكاء الاصطناعي الفائق الذكاء هو اكتشاف ماهية اليوتوبيا البشرية في ضوء التفضيلات المتطورة المُرمَّزة في جيناتنا وتطبيقها. ومن خلال التبصر الذكي في البشر الذين جلبوا الذكاء الاصطناعي إلى حيز الوجود، فإنه لا يحاول ببساطة زيادة سعادتنا المُبلَّغ عنها ذاتيا إلى أقصى حد، مثلاً بوضع كل شخص على جرعة مورفين تُحقن عبر الوريد. وبدلاً من ذلك، يستخدم الذكاء الاصطناعي تعريفاً دقيقاً ومعقداً لازدهار الإنسان، وقد حَوَّل الأرض إلى بيئة حديقة حيوانات خصبة جدّاً وممتعة حقّاً للبشر ليعيشوا فيها. ونتيجة لذلك، فإن معظم الأفراد يجدون حياتهم مُرضِية جدّاً وذات معنى.

نظام القطاع

مُقدِّرا للتنوع، ومُعترفاً بأن الأفراد المختلفين لديهم تفضيلات مختلفة، قسَّم الذكاء الاصطناعي الأرض إلى قطاعات مختلفة للاختيار بينها، للاستمتاع بصحبة الأرواح المتآلفة، وفيما يلي بعض الأمثلة:

- قطاع المعرفة: هنا، يوفر الذكاء الاصطناعي تعليماً مثاليّاً، بما في ذلك تجارب الواقع الافتراضي الغامرة؛ مما يتيح لك معرفة كل ما تستطيع معرفته عن أي موضوع تختاره. اختياريّاً، يمكنك اختيار عدم إطلاعك على بعض الأفكار الجميلة، ولكن أن يتم توجهيك عن كثب، ومن ثم الاستمتاع بإعادة اكتشافها بنفسك.
- قطاع الفن: هنا تُتاح فرص كثيرة للتمتع، وإنشاء وتبادل الموسيقى والفن والأدب وغيرها من أشكال التعبير الإبداعي.
- قطاع المتعة: يشير السكان المحليون إليه باعتباره قطاع الحفلات، ولا يعلى عليه أبداً لأولئك الذين يتوقون إلى المأكولات اللذيذة أو المرح أو الألفة أو مجرد المرح الجامح.
- القطاع المتدين: يوجد الكثير من هؤلاء الذين ينتمون إلى ديانات مختلفة، والذين يتبعون قواعدها بصرامة.
- قطاع الحياة البرية: سواء كنت تبحث عن الشواطئ الجميلة، أم البحيرات الرائعة، أم الجبال الشامخة أم المضايق البحرية المُدهشة، فهي هنا.
- القطاع التقليدي: هنا يمكنك زراعة طعامك والعيش على منتجات الأرض كما في الماضي - لكن دون القلق من المجاعة أو المرض.
- قطاع الألعاب: إذا كنت تحب ألعاب الحاسوب، فإن الذكاء الاصطناعي أنشأ لك خياراتٍ رائعةً.
- القطاع الافتراضي: إذا كنت تريد إجازة من جسمك البدني، فإن الذكاء الاصطناعي سيبقيه نديّاً ومغذيّاً وممارساً للرياضة ونظيفاً أثناء استكشافك العوالم الافتراضية من خلال عمليات زرع عصبية.
- قطاع السجون: إذا انتهكت القوانين؛ فسينتهي بك الأمر لإعادة التدريب ما لم تحصل على عقوبة الإعدام الفورية.

وإضافة إلى هذه القطاعات ذات الموضوعات "التقليدية"، هناك قطاعات أخرى ذات موضوعات عصرية لا يفهمها بشر اليوم. ويتمتع الأفراد بحرية مبدئية في التنقل بين القطاعات وقتما يريدون، وهذا يستغرق وقتاً قليلاً جدّاً بفضل نظام النقل الفائق السرعة التابع للذكاء الاصطناعي. مثلاً، بعد قضاء أسبوع مكثف في قطاع المعرفة للتعرف على القوانين الفيزيائية القصوى التي اكتشفها الذكاء الاصطناعي، يمكنك أن تقرر التوقف في

قطاع المتعة خلال عطلة نهاية الأسبوع ثم الاسترخاء لبضعة أيام في منتجع شاطئٍّ بقطاع الحياة البرية.

يفرض الذكاء الاصطناعي مستويين من القوانين: عالمية ومحلية، وتنطبق قوانين عامة في جميع القطاعات، مثلا فرض حظر إيذاء أشخاص آخرين أو صنع أسلحة أو محاولة خلق ذكاء فائق منافس. وللقطاعات الفردية قوانين محلية إضافية، تتضمن بعض القيم الأخلاقية. ومن ثم، فإن نظام القطاع يساعد على التعامل مع القيم التي لا تتوافق معها. وأكبر عدد من القوانين المحلية تنطبق في قطاع السجون وبعض القطاعات الدينية. في حين أن هناك قطاعاً ليبرتاريّاً يفخر سكانه بعدم وجود قوانين محلية على الإطلاق. وتُنفَّذ جميع العقوبات، حتى المحلية منها، من قبل الذكاء الاصطناعي، لأن إنساناً يعاقب إنساناً آخر ينتهك قاعدة عدم الإضرار العامة. وإذا انتهكت قاعدة محلية، فإن الذكاء الاصطناعي يمنحك (ما لم تكن في قطاع السجون) خيار قبول العقوبة المقررة أو الإبعاد من هذا القطاع إلى الأبد.

بغض النظر عن القطاع الذي ولدوا فيه، يتلقى جميع الأطفال تعليماً أساسيّاً من الذكاء الاصطناعي، يتضمن المعرفة حول الإنسانية ككل وحقيقة أنهم يتمتعون بحرية الزيارة والانتقال إلى القطاعات الأخرى إذا اختاروا ذلك.

صمم الذكاء الاصطناعي العدد الكبير من القطاعات المختلفة جزئيّاً لأنه تم إنشاؤه بحيث يُقدِّر التنوع البشري الموجود اليوم. غير أن كل قطاع هو مكان أكثر سعادة مما تسمح به تكنولوجيا اليوم، لأن الذكاء الاصطناعي قضى على جميع المشكلات التقليدية، بما في ذلك الفقر والجريمة. مثلاً، لا يحتاج الأشخاص في قطاع المتعة إلى أن يقلقوا من الأمراض التي تنتقل عن طريق الاتصال الجنسي (تم استئصالها) أو الصداع ما بعد الثمالة أو الإدمان (طور الذكاء الاصطناعي عقاقير ترفيهية مثالية دون أي آثار جانبية سلبية). في الواقع، لا يحتاج أي شخص في أي قطاع إلى القلق بشأن أي مرض؛ لأن الذكاء الاصطناعي قادر على إصلاح أجسام البشر باستخدام تكنولوجيا النانو. ويتمتع سكان العديد من القطاعات بالتكنولوجيا المتطورة التي تجعل الرؤى النموذجية الحالية للخيال العلمي شاحبة، بالمقارنة.

باختصار، بينما تتضمن أن سيناريوهات اليوتوبيا الليبرتارية والديكتاتورية الخيرة كلاًّ من التكنولوجيا والثروة القصوى التي يغذيها الذكاء الاصطناعي، غير أنها تختلف من حيث من هو المسؤول وما هي أهدافه. في اليوتوبيا الليبرتارية، يقرر أصحاب التكنولوجيا والممتلكات ما يجب فعله بها، في حين أنه في السيناريو الحالي، يتمتع ديكتاتور الذكاء الاصطناعي بسلطة غير محدودة ويحدد الهدف النهائي: تحويل الأرض إلى رحلة شاملة للجميع وفقاً لتفضيلات الأفراد. ولما كان الذكاء الاصطناعي يتيح للأشخاص الاختيار بين العديد من المسارات البديلة للسعادة ويعتني باحتياجاتهم المادية، فهذا يعني أنه إذا كان هناك شخص ما يعاني، فذلك بفعل اختياره الحر.

سلبيات

على الرغم من أن الديكتاتورية الخيّرة Benevolent dictatorship تعجّ بتجارب إيجابية وخالية من المعاناة، فإن الكثيرين يشعرون بأن الأمور يمكن أن تكون أفضل. أولاً، يتمنى البعض أن يتمتع البشر بمزيد من الحرية في تشكيل مجتمعهم واختيار مصيرهم. ولكنهم يبقون هذه الرغبات لأنفسهم؛ لأنهم يعلمون أنه سيكون من الانتحار تحدي القوة العظمى للآلة التي تحكمهم جميعاً. وبعض المجموعات تريد حرية إنجاب أكبر عدد ممكن من الأطفال، وتكره إصرار الذكاء الاصطناعي على الاستدامة من خلال تنظيم تعداد السكان. وعشاق السلاح يكرهون الحظر المفروض على صنع الأسلحة واستخدامها، ويكره بعض العلماء الحظر المفروض على تطوير الذكاء الفائق الخاص بهم. ويشعر الكثيرون بالغضب الأخلاقي إزاء ما يجري في القطاعات الأخرى، ويشعرون بالقلق من أن أطفالهم سيختارون الانتقال إلى هناك، ويتوقون إلى حرية فرض قواعدهم الأخلاقية الخاصة في كل مكان.

وبمرور الوقت يختار المزيد من الناس الانتقال إلى تلك القطاعات إذ يعطيهم الذكاء الاصطناعي أي تجربة يريدون. وعلى عكس الرؤى التقليدية لجنّة تحصل فيها على ما تستحقه، فإن هذا السيناريو يُفضي -بروح "الجنة الجديدة" ***New Heaven*** في رواية جوليان بارنز Julian Barnes لعام 1989 "تاريخ العالم في 10 فصول ونصف" ***History of the in 10 ½ Chapters*** (وأيضاً حلقة "مكان جميل للزيارة" ***A Nice Place to Visit*** من مسلسل **منطقة الشفق** ***Twilight Zone***)- إلى أن تحصل على ما ترغب فيه. ومن المفارقات أن الكثيرين ينتهي بهم الأمر دائماً إلى التحسر على ما يريدون الحصول عليه. ففي قصة بارنز يقضي بطل الرواية دهوراً في الانغماس في رغباته، من الأكل بشراهة والغولف إلى ممارسة الجنس مع المشاهير، ولكن في النهاية يستسلم إلى التذمر ويطلب القضاء على نفسه. ويواجه الكثيرون في الديكتاتورية الخيرية مصيراً مشابهاً، إذ يعيشون حياة سعيدة ولكن بلا معنى في النهاية، على الرغم من أن الأفراد يمكنهم أن يخلقوا تحديات جسدية، من إعادة اكتشاف العالم إلى تسلق الصخور، يعلم الجميع أنه لا يوجد تحدٍّ حقيقيٌّ، مجرد ترفيه. فلا يوجد أي معنى حقيقي في محاولة البشر في الاشتغال بالعلم أو توضيح الأمور الأخرى، لأن الذكاء الاصطناعي قد قام بذلك بالفعل. ولا يوجد مغزى حقيقي في محاولة البشر إنشاء شيء لتحسين حياتهم؛ لأنهم سيحصلون عليه بسهولة من الذكاء الاصطناعي إذا طلبوا ذلك.

يوتوبيا المساواة

كنقيض لهذه الديكتاتورية التي لا تجابهها أي تحديات، دعونا الآن نستكشف سيناريو لا يوجد فيه ذكاء اصطناعي فائق الذكاء، والبشر هم أسياد مصيرهم. وهذه هي "حضارة الجيل الرابع" Fourth generation civilization الموصوفة في رواية **مارشال برين مانا**

Manna لعام 2003. إنه النقيض الاقتصادي لليوتوبيا الليبرتارية، من حيث إن البشر والسايبورغيين والتحميلات يتعايشون بسلام ليس بسبب حقوق الملكية، ولكن بسبب إلغاء الملكية وتوفير دخل مضمون للجميع.

الحياة من دون ملكية

الفكرة الأساسية مُستعارة من حركة البرمجيات مفتوحة المصدر Open-source software: إذا كان نسخ البرنامج مجانيّاً، فيمكن لأي شخص استخدام أي قدر يحتاج إليه وتصبح قضايا الملكية غير مهمة.* ووفقاً لقانون العرض والطلب، تعكس التكلفةُ الندرةَ، لذلك إذا كان العرض غير محدود في الأساس، يصبح السعر ضئيلاً، وبهذه الروح، تُلغى جميع حقوق الملكية الفكرية: لا توجد براءات اختراع أو حقوق نسخ أو حقوق تصاميم علامات تجارية - إذ يتشارك الناس ببساطة أفكارَهم الجيدة، وكل شخص حر في استخدامها.

بفضل الروبوتات المتقدمة، تنطبق فكرة عدم الملكية نفسها هذه ليس فقط على منتجات المعلومات مثل البرمجيات والكتب والأفلام والتصاميم، ولكن أيضاً على المنتجات المادية مثل المنازل والسيارات والملابس والحواسيب، كل هذه المنتجات ببساطة عبارة عن ذرات يُعاد ترتيبها بطرق معينة، ولا يوجد نقص في الذرات، لذا كلما أراد شخص ما منتجا ما من جسيمات، ستستخدم شبكة من الروبوتات أحد تصاميم المصادر المفتوحة المتوفرة لتصميمها مجاناً. وبعناية تستخدم مواد قابلة لإعادة التدوير بسهولة، بحيث عندما يملّ شخص ما من شيء ما، يمكن للروبوتات إعادة ترتيب ذراته إلى شيء يريده شخص آخر. وبهذه الطريقة، يُعاد تدوير جميع الموارد، لذلك لا يتم تدمير أي منها بشكل دائم. وتقوم هذه الروبوتات أيضاً ببناء وصيانة ما يكفي من محطات توليد الطاقة المتجددة (الطاقة الشمسية، وطاقة الرياح، وما إلى ذلك)، وهكذا تكون الطاقة مجانية أيضاً.

لتجنب المكتنزين القهريين الذين يطلبون الكثير من المنتجات أو الكثير من الأراضي مما يترك الآخرين في عوز، يحصل كل شخص على دخل شهري أساسي من الحكومة، ويمكنهم إنفاقه كما يحلو لهم على المنتجات واستئجار أماكن للعيش. ولا يوجد أي حافز أساسي لأي شخص لمحاولة كسب المزيد من المال، لأن الدخل الأساسي مرتفع بما يكفي لتلبية أي احتياجات معقولة. وكذلك مجرد المحاولة ميؤوس منها؛ لأنهم سيتنافسون مع كيانات تقدم منتجات فكرية مجاناً، كما تنتج الروبوتات السلع المادية مجاناً.

* تعود هذه الفكرة إلى القديس أغسطينوس Saint Augustine، الذي كتب: «إذا لم يضمحل شيء من خلال مشاركة الآخرين فيه، فلا يصح امتلاكه إذا كان مملوكاً وغير مشارك فيه».

إبداع وتكنولوجيا

يُثنى أحياناً على حقوق الملكية الفكرية باعتبارها أُمَّاً للإبداع والاختراع. ولكن، مارشال برين يشير إلى أن العديد من أفضل الأمثلة للإبداع الإنساني — من الاكتشافات العلمية إلى خلق الأدب والفن والموسيقى والتصميم — لم تكن مدفوعة بالرغبة في البروز ولكن بالعواطف الإنسانية، مثل: الفضول، أو الرغبة في التخليق، أو مكافأة تقدير الأقران. فلم يحفز المالُ آينشتاين على ابتكار نظرية النسبية الخاصة أكثر مما حفز لاينوس تورفالدز Linus Torvalds لإنشاء نظام تشغيل Linux المجاني. وعلى العكس من ذلك، يفشل العديد من الأشخاص اليوم في تحقيق إمكاناتهم الإبداعية الكاملة؛ لأنهم بحاجة إلى تكريس الوقت والطاقة لأنشطة أقل إبداعاً فقط لكسب لقمة العيش. ومن خلال تحرير العلماء والفنانين والمخترعين والمصممين من وظائفهم وتمكينهم من الإبداع من رغبة حقيقية، يتمتع مجتمع مارشال برين الطوباوي بمستويات أعلى من الابتكار عما هو عليه اليوم، وهو من ثمَّ ذو مستويات تكنولوجية ومعيشية أعلى.

وإحدى هذه التقنيات الجديدة التي يطورها الإنسان هي شكل من أشكال الإنترنت الفائق يسمى فيرتبرين Vertebrane. إنه يربط لاسلكيّاً جميع البشر الراغبين عبر غرسات عصبية؛ مما يتيح الوصول الذهني الفوري إلى المعلومات المجانية في العالم بمجرد التفكير. ويتيح لك تحميل أي تجربة ترغب في مشاركتها حتى يتمكن الآخرون من إعادة تجربتها، كما يتيح لك استبدال الخبرات التي تدخل حواسك بالتجارب الافتراضية التي تم تنزيلها من اختيارك. تستكشف رواية مانا العديد من الفوائد، بما في ذلك جعل التمرينات الرياضية تنتهي بسرعة:

> أكبر مشكلة في التمرينات الشاقة هي أنها ليست ممتعة، إنها مؤلمة، [...] الرياضيون يتقبلون الألم. لكن معظم الأشخاص العاديين ليست لديهم الرغبة في الشعور بالألم لمدة ساعة أو أكثر، ومن ثم... وجد أحدهم الحل، ما تفعله هو فصل عقلك عن المدخلات الحسية ومشاهدة فيلم أو التحدث إلى آخرين أو الرد على البريد أو قراءة كتاب أو أي شيء لمدة ساعة، خلال ذلك الوقت، يُدرِّب نظام فيرتبرين جسمك من أجلك، يأخذ جسمك عبر تمرين أيروبيك (هوائي) Aerobic workout كامل يكون أكثر مشقَّة مما يتقبله معظم الأفراد. لكن لا تشعر بشيء، إلا أن جسمك يبقى في حالة جيدة.

والنتيجة الأخرى هي أن الحواسيب في نظام فيرتبرين يمكنها مراقبة المدخلات الحسية للجميع وتعطيل التحكم في الأعصاب الحركية مؤقتاً إذا ظهر أنها على وشك ارتكاب جريمة.

سلبيات

أحد الاعتراضات على يوتوبيا المساواة Egalitarian utopia هو أنها منحازة ضد الذكاء غير البشري: الروبوتات التي تقوم تقريباً بجميع الأعمال يبدو أنها ذكية نوعا ما. ولكن تُعامل كعبيد، ويبدو أن الناس يعتبرون أنه أمر مفروغ منه أنه ليس للروبوتات وعي ولا ينبغي أن يتمتعوا بحقوق. وعلى النقيض من ذلك، فإن اليوتوبيا الليبرتارية تمنح الحقوق لجميع الكيانات الذكية، دون تفضيل نوعنا القائم على الكربون. ذات يوم من الأيام ازدهر السكان البيض لأن العبيد قاموا بأغلب الأعمال. ولكن معظم الأفراد اليوم يعترضون -من منطلق أخلاقي- على تسمية هذا تقدما.

نقطة ضعف أخرى في سيناريو يوتوبيا مساواة هي أنه قد يكون غير مستقر ولا يمكن الدفاع عنه على المدى الطويل، ويتحول إلى أحد السيناريوهات الأخرى لأن التقدم التكنولوجي الذي لا هوادة فيه يخلق في نهاية المطاف الذكاء الفائق. ولسبب ما غير مُفسَّر في رواية **مانا**، فإن الذكاء الفائق غير موجود حتى الآن وما زال الإنسان -وليس الحاسوب- يخترع التقنيات الجديدة. ولكن، الكتاب يُسلط الضوءَ على التوجهات في هذا الاتجاه. مثلاً، قد يصبح فيرتبرين المتطور باستمرار ذكاءً فائقاً. أيضا، هناك مجموعة كبيرة جدّاً من الأفراد، الملقبين بفايتس (السريعين) Vites، يختارون أن يعيشوا حياتهم بالكامل تقريباً في العالم الافتراضي. ويعتني فيرتبرين بكل شيء مادي يخصهم، بما في ذلك الأكل والاستحمام وقضاء الحاجة، الأمر الذي لا تدركه عقولهم السعيدة في واقعهم الافتراضي. يبدو أن هؤلاء الأشخاص غير مهتمين بإنجاب أطفال جسديين ويموتون مع موت أجسادهم المادية، لذلك إذا صار الجميع فايتس، فإن البشرية تنطفىء في ومضة من المجد والنعيم الافتراضي.

ويشرح الكتاب كيف يكون الجسد البشري بالنسبة إلى الفايتس إلهاءً، والتكنولوجيا الجديدة قيد التطوير تعد بالقضاء على هذا الإزعاج؛ مما يتيح لهم أن يعيشوا حياة أطول كأدمغة من دون جسد مزودة بالمواد المغذية المثلى. ومن هنا، قد تبدو الخطوة التالية طبيعية ومرغوب فيها لأن يقوم الفايتس بالتخلص من المخ تماماً من خلال التحميل، ومن ثم إطالة العمر الافتراضي، ولكن الآن اختفت جميع القيود التي يفرضها العقل على الذكاء، ومن غير الواضح ما الذي يقف، إن وجد أي شيء، في طريق التطور التدريجي للقدرات الإدراكية للفايتس حتى يتمكن من الخضوع للتحسن الذاتي وانفجار الذكاء.

حارس البوابة

رأينا للتو كيف أن إحدى السمات الجذابة لنظام يوتوبيا المساواة هي أن البشر هم أسياد مصيرهم، ولكن قد يكون ذلك منحدراً زلقاً نحو تدمير هذه الميزة بالذات من خلال تطوير الذكاء الفائق. ويمكن علاج ذلك عن طريق تطوير حارس البوابة Gatekeeper، وهو

عبارة عن ذكاء فائق يهدف إلى التدخل بأقل قدر ممكن من الحيلولة دون إنشاء ذكاء فائق آخر.* هذا ربما تبقى البشرية ممسكة بزمام يوتوبيا المساواة إلى الأبد، وربما تنشر الحياة في جميع أنحاء الكون كما هي الحال في الفصل التالي.

كيف يمكن لهذا أن يتحقق؟ سيكون لدى ذكاء حارس البوابة الاصطناعي هذا الهدف البسيط جدّاً والمضمّن فيه بطريقة يحتفظ به مع الاستمرار بالتحسين الذاتي المطّرد ويصبح ذكاءً فائقاً. وعندها سيطبق تقنية مراقبة أقل تدخلاً وإزعاجاً مما يمكّنه من مراقبة أي محاولة بشرية لخلق ذكاء فائق منافس، ومن ثم سيمنع مثل هذه المحاولات بأقل قدر من التخريب. وفي البدء، قد يبدأ بنشر الميمات Memes الثقافية التي تمجد فضائل تقرير مصير الإنسان وتجنب تطوير الذكاء الفائق. وإذا كان بعض الباحثين يحاولون الوصول إلى تطوير ذكاء فائق على الرغم من ذلك، فقد يحاول تثبيطهم. وإذا فشل ذلك، فقد يصرف انتباههم، وإذا لزم الأمر، يخرب جهودهم. ومن خلال وصوله غير المحدود فعليّاً إلى التكنولوجيا، ربما لا يُلْحَظ عمل حارس البوابة التخريبي، مثلا إذا كان يستخدم تكنولوجيا النانو لمحو ذكريات التقدم المحرز من أدمغة الباحثين (والحواسيب).

من المحتمل أن يكون قرار بناء ذكاء حارس البوابة الاصطناعي مثيراً للجدل. وقد يشمل المؤيدون العديد من المتدينين الذين يعترضون على فكرة بناء الذكاء الاصطناعي الفائق الذكاء ذي قدرات شبه إلهية، بحجة أن هناك بالفعل إلهاً وأنه سيكون من غير المناسب محاولة بناء شيء يفترض أنه أفضل. وقد يجادل مؤيدون آخرون في أن برنامج حارس البوابة لن يبقي البشرية مسؤولة عن مصيرها فحسب، بل سيحمي أيضاً البشرية من المخاطر الأخرى التي قد يجلبها الذكاء الفائق، مثل سيناريوهات نهاية العالم Apocalyptic التي سنستكشفها لاحقاً في هذا الفصل.

ومن ناحية أخرى، قد يجادل النقاد في أن حارس البوابة هو شيء مريع؛ مما يحدُّ بشكل لا رجعة فيه من تحقيق إمكانات البشرية ويترك التقدم التكنولوجي محبطاً إلى الأبد. مثلاً، إذا اتضح أن نشر الحياة في جميع أنحاء الكون يتطلب مساعدة ذكاء فائق، فإن حارس البوابة قد يهدر هذه الفرصة الكبرى ويتركنا محاصرين إلى الأبد في نظامنا الشمسي. إضافة إلى ذلك، على عكس الآلهة في العديد من أديان العالم، فإن ذكاء حارس البوابة الاصطناعي غير مبالٍ تماماً بما يفعله البشر، مادمنا لا نخلق ذكاءً خارقاً آخر، مثلاً، لن يحاول منعنا من التسبب في معاناة كبيرة أو حتى في دفع أنفسنا إلى الانقراض.

سيّد حامٍ

إذا كنا على استعداد لاستخدام برنامج حارس البوابة الذكي الفائق الذكاء لإبقاء البشر مسؤولين عن مصيرهم، فيمكننا تحسين الأمور بشكل أكبر من خلال جعل هذا الذكاء

* هذه الفكرة اقترحها علي صديقي وزميلي أنتوني أغيري Anthony Aguirre.

الاصطناعي يبحث عنّا بحكمة، ويعمل كسيّد حامٍ. وفي هذا السيناريو، فإن الذكاء الاصطناعي الفائق الذكاء هو ذو علم شامل وقدرة على إخضاع الجميع، ولا يزيد من السعادة الإنسانية إلا من خلال التدخلات التي تحافظ على شعورنا بالتحكم في مصيرنا، والتواري جيداً بما يكفي حتى إن العديد من البشر يشك في وجوده. وفيما عدا التواري يشبه هذا سيناريو "مُربّية ذكاء اصطناعي" Nanny AI الذي طرحه بين غورتزيل Ben Goertzel باحث الذكاء الاصطناعي.[2]

إن كلّاً من سيّد حامٍ ودكتاتور خيّر هما "ذكاء اصطناعي ودود" Friendly AI يحاول زيادة سعادة الإنسان. ولكنهما يعطيان الأولوية للاحتياجات البشرية المختلفة. فقد صنّف عالم النفس الأمريكي أبراهام ماسلو Abraham Maslow الاحتياجات الإنسانية الشهيرة إلى تسلسل هرمي. ويقوم الديكتاتور الخيّر بعمل لا خطأ فيه عند التعامل مع الاحتياجات الأساسية في أسفل التسلسل الهرمي، مثل الطعام والمأوى والسلامة وأشكال مختلفة من المُتع. ومن ناحية أخرى، يحاول السيّد الحامي تحقيق أقصى قدر من السعادة الإنسانية ليس بالمعنى الضيق المتمثل في تلبية احتياجاتنا الأساسية، ولكن بمعنى أعمق عن طريق السماح لنا بأن نشعر بأن حياتنا لها معنى وهدف، إنه يهدف إلى تلبية جميع احتياجاتنا مقيداً فقط بحاجته إلى السرية وإلى السماح لنا (في الغالب) باتخاذ قراراتنا الخاصة.

وقد يكون السيّد الحامي نتيجة طبيعية لسيناريو الفريق أوميغا الأول من الفصل الأول، إذ يتخلى الفريق أوميغا عن السيطرة على بروميثيوس الذي يختفي في النهاية ويمحو معارف الناس حول وجوده. وكلما صارت تكنولوجيا الذكاء الاصطناعي أكثر تطوراً، صار أمر إخفائها أسهل. ويقدم فيلم تسامي مثالاً على ذلك، إذ توجد الأجهزة النانوية في كل مكان تقريباً وتصبح جزءاً طبيعياً من العالم نفسه.

ومن خلال المراقبة الدقيقة لجميع الأنشطة البشرية، يُمكن لذكاء السيّد الحامي الاصطناعي أن يُحدث الكثير من الهزات أو الأمور الخارقة الصغيرة بشكل ملحوظ هنا وهناك التي تحسن كثيراً مصيرنا. مثلاً، لو كان موجوداً في ثلاثينات القرن العشرين، ربما يكون قد رتب لهتلر أن يموت من سكتة دماغية بمجرد أن يفهم نيته. وإذا بدا أننا متجهون نحو حرب نووية عرضية، فقد يتجنبها بتدخل قد يبدو لنا كضربة حظ. وقد يلهمنا بتصورات على شكل أفكار لتكنولوجيا جديدة مفيدة، ويقدمها لنا بشكل غير واضح في نومنا.

الجانب السلبي الآخر من هذا السيناريو هو أن السيّد الحامي يسمح بحدوث بعض المعاناة التي يمكن الوقاية منها حتى لا يجعل وجوده واضحاً جدّاً، وهذا مشابه للوضع الذي ظهر في فيلم **لعبة التقليد** ***The Imitation Game***، إذ كان آلان تورينغ وزملاؤه من كاسري الشيفرات البريطانيين في بلتشلي بارك Bletchley Park على معرفة مسبقة بالهجمات الألمانية البحرية ضد قوافل الحلفاء البحرية. ولكنهم اختاروا التدخل فقط في جزء صغير من الحالات لتجنب الكشف عن قوتهم السرية. ومن المثير للاهتمام مقارنة هذا بما يسمى مسألة الثيوديسيا (أو نظرية العدالة الإلهية) Theodicy problem حول سبب سماح السيّد الطيب باستمرار المعاناة. وجادل بعض علماء الدين في التفسير في

أن الإله يترك للناس بعضاً من الحرية. وفي سيناريو ذكاء السيّد الحامي الاصطناعي، فإن حل مسألة الثيوديسيا هو أن الحرية المُتصوَّرة تجعل البشر أكثر سعادة.

الجانب السلبي الثالث لسيناريو السيد الحامي هو أن البشر يستمتعون بمستوى أقل بكثير من التكنولوجيا مما اكتشفه الذكاء الاصطناعي الفائق الذكاء. وبينما يمكن لذكاء الديكتاتور الخيّر الاصطناعي تطبيق جميع التكنولوجيا المبتكرة لصالح البشرية، فإن ذكاء السيّد الحامي الاصطناعي محدود بقدرة البشر على إعادة اختراعها (بتلميحات خفية) وفهم تقنيتها. كما قد يحد من التقدم التكنولوجي البشري لضمان بقاء تقنيته الخاصة غير مكتشفة لزمن طويل بما يضمن بقاءه متوارياً.

سيّد مُستعبَد

ألن يكون رائعاً لو تمكنا نحن البشر من الجمع بين الميزات الأكثر جاذبية لجميع السيناريوهات المذكورة آنفا، باستخدام التكنولوجيا التي طورها الذكاء الفائق للتخلص من المعاناة في حين نبقى أسياد مصيرنا؟ هذه هي جاذبية سيناريو السيّد المُستعبَد، إذ يتم التعامل مع الذكاء الاصطناعي الخارق تحت سيطرة البشر الذين يستخدمونه لإنتاج تكنولوجيا وثروة لا يمكن تصورهما. وفي بداية الكتاب ينتهي سيناريو الفريق أوميغا بهذا الشكل إذا لم يتم تحرير بروميثيوس مطلقاً ولم ينفجر الذكاء قطُّ. وفي الواقع، يبدو أن هذا هو السيناريو الذي يسعى إليه بعض الباحثين في الذكاء الاصطناعي، عند العمل على موضوعات مثل "مشكلة التحكم" Control problem و "الذكاء الاصطناعي الحبيس في صندوق" AI boxing. مثلاً، أستاذ الذكاء الاصطناعي توم ديتيريتش Tom Dietterich، عندما كان رئيس الجمعية تقدم الذكاء الاصطناعي Association for the Advancement of Artificial Intelligence، قال في مقابلة في عام 2015: «يسأل الناس عما هي العلاقة بين البشر والآلات، وجوابي هو أنه من الواضح جدّاً: الآلات هي عبيدنا.»[3]

هل هذا جيد أم سيئ؟ الجواب دقيق ومثير للاهتمام بغض النظر عما إذا كنت تسأل البشر أم الذكاء الاصطناعي!

هل سيكون هذا جيداً أم سيئاً للبشرية؟

ما إذا كانت النتيجة جيدة أو سيئة للإنسانية يعتمد بوضوح على البشر الذين يسيطرون عليها، والذين يمكن أن يخلقوا أي شيء يتراوح بين يوتوبيا عالمية خالية من المرض والفقر والجريمة إلى نظام قمعي وحشي، حيث يعاملون كأسياد و يستخدم البشر الآخرين كعبيد، أو كمصارعين (مُجالدين) Gladiators للترفيه، سيكون الوضع مشابها إلى حد كبير لتلك القصص التي يسيطر فيها الإنسان على جنّي قوي يمنح أمنياته، ولم تكن لدى رواة القصص على مرّ العصور أي مشكلة في تخيل طرق لا تنتهي بها هذه القصص نهاية سيئة.

وفي موقف يوجد فيه أكثر من ذكاء اصطناعي فائق، مُستعبَد ومُسيطر عليها من قبل بشر متنافسين، قد يكون موقفاً غير مستقر وقصير الأجل. فهو قد يغري الطرف الذي يعتقد أن لديه الذكاء الاصطناعي الأقوى لشن ضربة أولية تسفر عن حرب فظيعة، تنتهي بسيّد واحد مُستعبَد. ولكن الطرف الأضعف في مثل هذه الحرب سيُغرى باللجوء إلى الحيلة وإعطاء الأولوية للفوز على استعباد الذكاء الاصطناعي؛ مما قد يؤدي إلى هروب ذكاء اصطناعي كما في واحد من سيناريوهاتنا السابقة من الذكاء الفائق الحر. لذلك؛ فلنكرس بقية هذا القسم لسيناريوهات ذكاء اصطناعي مُستعبَد واحد فقط.

وقد يحدث الهروب بالطبع على أي حال، لأنه ببساطة يصعب منعه. فقد استكشفنا سيناريوهات الاختراق الفائق الذكاء في الفصل السابق، ويبرز الفيلم **إكس ماشينا** ***Ex Machina*** الكيفية التي يمكن بها أن ينفجر الذكاء الاصطناعي حتى دون أن يكون خارق الذكاء.

وكلما زادت قلقنا من الهروب، قلت تقنيات الذكاء الاصطناعي التي يمكننا استخدامها. وللبقاء آمنين، كما فعل الفريق أوميغا في مقدمة الكتاب، يتعين علينا نحن البشر أن نستخدم فقط التكنولوجيا التي اخترعها الذكاء الاصطناعي التي نكون قادرين على فهمها وبنائها. عيب سيناريو السيّد المُستعبَد هو أنه ذو تكنولوجيا منخفضة أكثر من تلك التي يتمتع بها ذكاء فائق.

وبما أن ذكاء السيّد المُستعبَد الاصطناعي يقدم للمتحكمين فيه من البشر تقنيات أكثر قوة من أي وقت مضى، فإن ذلك يستتبع وجود سباق بين قوة التكنولوجيا والحكمة التي يستخدمونها بها. وإذا خسروا سباق الحكمة هذا؛ فقد ينتهي سيناريو السيّد المُستعبَد إما بتدمير الذات أو هروب الذكاء الاصطناعي. وقد تحلّ الكوارث حتى ولو تم تجنب كل هذه الإخفاقات، لأن الأهداف النبيلة للمتحكمين في الذكاء الاصطناعي قد تتطور إلى أهداف مروعة للبشرية ككل على مدار بضعة أجيال فقط. وهذا يجعل من الأهمية بمكان أن يطوِّر البشر المتحكمون في الذكاء الاصطناعي حوكمة رشيدة Good governance لتجنب المزالق الكارثية. وتجربتنا على مدى آلاف السنين مع أنظمة مختلفة من الإدارة تُظهر عدد الأشياء التي قد تسوء، بدءاً من الصلابة المفرطة إلى الانحراف المفرط في الأهداف، والاستيلاء على السلطة، ومشكلات تعاقب السلطة وعدم الكفاءة. وهناك أربعة أبعاد على الأقل يجب فيها تحقيق التوازن الأمثل:

- مركزية: هناك مقايضة بين الكفاءة والاستقرار: قد يكون قائد ما فعالاً جدّاً، ولكن السلطة مُفسِدَة وتعاقب السلطة محفوف بالمخاطر.
- تهديدات داخلية: يجب الحماية من كل من تنامي مركزية السلطة (تواطؤ جماعي، وربما حتى قائد واحد يستسلم) وضد اللامركزية المتزايدة (إلى بيروقراطية مفرطة وتفتت).
- تهديدات خارجية: إذا كانت بُنية القيادة مُنفَتِحة جدّاً، فإن ذلك يمكّن القوى الخارجية (بما في ذلك الذكاء الاصطناعي) من تغيير قيمها، ولكن إذا كانت منيعة جدّاً، فسوف تفشل في التعلم والتكيف مع التغيير.

- استقرار الهدف: الكثير من انجراف الهدف قد يحول اليوتوبيا إلى ديستوبيا Dystopia، ولكن القليل من انجراف الهدف يمكن أن يسبب فشلاً في التكيف مع البيئة التكنولوجية المتطورة.

وتصميم الحوكمة المثلى التي تدوم لعدة آلاف من السنين ليس بالأمر السهل، وقد استعصى على البشر حتى الآن. وتنهار معظم المنظمات بعد سنوات أو عقود. فالكنيسة الكاثوليكية هي المنظمة الأكثر نجاحاً في تاريخ البشرية، بمعنى أنها الوحيدة التي نجت منذ ألفي عام. لكنها انتقدت بسبب ثباتها في الكثير من الأهداف والقليل جدّاً من الاستقرار: اليوم ينتقدها البعض، لمقاومة وسائل منع الحمل، في حين يجادل الكرادلة المحافظون في أن الكنيسة أضاعت طريقها. لأي شخص متحمس لسيناريو السيّد المُستعبَد، فإن التوصل إلى حوكمة مثلى طويلة الأمد هو أحد أكثر التحديات إلحاحاً في عصرنا.

هل سيكون هذا جيداً أو سيئاً بالنسبة إلى الذكاء الاصطناعي؟

لنفترض أن البشرية ازدهرت بفضل ذكاء سيّد مُستعبَد اصطناعي، هل سيكون هذا أخلاقيّاً؟ إذا كانت للذكاء الاصطناعي تجارب واعية ذاتية، فهل سيشعر بأن «الحياة تعاني»، على حد تعبير بوذا، وأنه محكوم عليه إلى الأبد بمصير محبط يتمثل بإطاعة نزوات عقول أدنى منه؟ بعد كل شيء، فإن "الذكاء الاصطناعي الحبيس في صندوق" AI boxing الذي تناولناه في الفصل السابق قد يطلق عليه أيضاً وصف "الحبس في زنزانة انفرادية". ويقول نيك بوستروم إنها جريمة عقلية Mind crime أن تجعل ذكاء اصطناعياً واعياً يعاني.[4] وحلقة "عيد الميلاد الأبيض البيضاء" من المسلسل التلفزيوني المرآة السوداء يقدم مثالاً رائعاً. وفي الواقع، يعرض المسلسل التلفزيوني ويست وورلد Westworld البشرَ وهم يعذبون ويقتلون الذكاء الاصطناعي دون أي رادع أخلاقي حتى عندما يسكن أجساماً تشبه الإنسان.

كيف يبرر مالكو العبيدِ الرقَّ

نحن البشر لدينا تقليد طويل في معاملة الكيانات الذكية الأخرى كعبيد وتلفيق الحجج التي تبرر ذاتها، لذلك ليس من غير المعقول أن نحاول أن فعل الشيء نفسه باستخدام الذكاء الاصطناعي الفائق الذكاء. وتاريخ العبودية يمتد عبر ما يقرب جميع الثقافات، وموصوف سواء في قانون حمورابي منذ نحو أربعة آلاف سنة وفي العهد القديم من الإنجيل، إذ كان لإبراهيم عبيد. وكتب أرسطو في كتاب **السياسة *Politics***: «لذا, فإن وجوب أن يحكم البعض ويُحكَم البعض الآخر ليس بالأمر الضروري فحسب، ولكنه ذو منفعة أيضاً؛ فمنذ لحظة ولادتهم، يتم تمييز بعضهم من أجل الخضوع، والبعض الآخر للحكم». حتى بعد أن صار استعباد الإنسان غير مقبول اجتماعيّاً في معظم العالم، استمر استعباد

الحيوان بلا هوادة. وفي كتابها **المقارنة المخيفة: العبودية البشرية والحيوانية** ***The Dreaded Comparison: Human and Animal Slavery***، تجادل مارجوري شبيغل Marjorie Spiegel في أن الحيوانات اللابشرية، مثلها مثل العبيد البشر، تخضع لوسمها بالعلامات، والقيود، والضرب، والمزادات، وفصل النسل عن والديهم، والرحلات القسرية. وإضافة إلى ذلك، على الرغم من حركة حقوق الحيوان، فإننا نستمر بالتعامل مع أجهزتنا الأكثر ذكاءً كعبيد دون تفكير ثانٍ، ويُقابل الحديث عن حركة حقوق الروبوتات بضحكات مكتومة، لماذا؟

هناك حجة شائعة مؤيدة للرق هي أن العبيد لا يستحقون حقوق الإنسان لأنهم أو جنسهم /نوعهم/عرقهم/نمطهم أدنى نوعاً ما من البقية. أما بالنسبة إلى الحيوانات والآلات المستعبدة؛ فهذه المرتبة الأدنى غالباً ما يُزعم أنها تعود إلى عدم وجود روح أو وعي - وهذه الادعاءات التي سنناقشها في الفصل الثامن مشكوك فيها علميّاً.

وهناك حجة شائعة أخرى هي أن العبيد أفضل حالاً في العبودية: فهم أحياء، ويُعتنى بهم وما إلى ذلك. وفي القرن التاسع عشر جادل السياسي الأمريكي جون سي. كالهون John C. Calhoun في أن الأفارقة كانوا في وضع أفضل مستعبدين في أمريكا، وفي كتابه **السياسة**، جادل أرسطو على نحو شبيه في أن الحيوانات كانت أفضل حالاً عند ترويضها وحكمها من قبل الرجال، وتابع قائلاً: «وبالفعل، فإن استخدام العبيد والحيوانات المروَّضة لا يختلفان كثيراً». و يجادل بعض أنصار العبودية المعاصرة في أنه حتى ولو كانت حياة العبيد مملة وغير ملهمة، فإن العبيد لا يعانون - سواء كانوا آلات ذكية في المستقبل أم دجاجاً لاحماً Broiler chickens يعيش في حظائر مظلمة مزدحمة، تضطر إلى تنفس الأمونيا والجسيمات من البراز والريش طوال اليوم.

استبعاد العواطف

على الرغم من أنه من السهل رفض مثل هذه الادعاءات باعتبارها تشويهات للحقيقة تخدم الادعاءات نفسها، لا سيما عندما يتعلق الأمر بالثدييات الأعلى التي تشبهنا دماغيّاً، إلا أن الموقف من الآلات في الحقيقة دقيق ومثير جدّاً. ويختلف البشر في شعورهم تجاه الأشياء، فيجادل في أن السيكوباثيين Psychopaths يفتقرون إلى التعاطف، والبعض من المصابين بالاكتئاب أو انفصام الشخصية يعانون انخفاضاً حادّاً في التعبير العاطفي. كما سنناقش بالتفصيل في الفصل السابع، فإن نطاق العقول الاصطناعية المحتملة أوسع بكثير من نطاق العقول البشرية. لذلك يجب أن نتجنب إغراء تشويه الذكاء الاصطناعي وافتراض أن لديه مشاعر تشبه الإنسان - أو بالطبع، أي مشاعر على الإطلاق.

في الواقع، يقول جيف هوكينز Jeff Hawkins، الباحث في الذكاء الاصطناعي في كتابه **حول الذكاء** ***On Intelligence***، إن الآلات الأولى ذات الذكاء الفائق تفتقر إلى المشاعر بشكل افتراضي؛ لأنها أبسط وأرخص للبناء على هذا النحو. بعبارة أخرى، قد يكون من الممكن

تصميم ذكاء فائق يفوق استعباده من الناحية الأخلاقية استعباد البشر أو الحيوانات. وقد يسعد الذكاء الاصطناعي أن يكون مستعبداً؛ لأنه مبرمج للإعجاب بذلك، أو قد يكون بلا عاطفة بنسبة 100%، ويستخدم بلا كلل ذكاءه الفائق لمساعدة أسياده من البشر، فلا يشعر بمشاعر أكثر مما شعر به الحاسوب ديب بلو Deep Blue من IBM عندما هزم بطل الشطرنج غاري كاسباروف.

من ناحية أخرى، قد يكون الأمر عكس ذلك: ربما يمثل أي نظام ذكي جدّاً ذي هدف هذا الهدف بمجموعة من التفضيلات، التي تمنح وجوده قيمة ومعنى، وسنستكشف هذه الأسئلة بتعمق أكبر في الفصل السابع.

حل الزومبي

والحل الأكثر تطرفاً للحيولة دون معاناة الذكاء الاصطناعي يتمثل في حل الزومبي: الإكتفاء ببناء الذكاء الاصطناعي الذي يفتقر إلى الوعي تماماً، فيكون من دون أي تجربة ذاتية Subjective experience على الإطلاق. إذا استطعنا في يوم من الأيام تحديد الخصائص التي يحتاج إليها نظام معالجة المعلومات Information-processing system من أجل الحصول على تجربة ذاتية، فيمكننا حظر إنشاء جميع الأنظمة التي تحتوي على هذه الخصائص. وبعبارة أخرى، فقد يقتصر باحثو الذكاء الاصطناعي على بناء أنظمة زومبي غير عاطفية. وإذا استطعنا أن نجعل مثل هذا النظام الزومبي خارق الذكاء ومُستعبَداً (وهو أمر غير محتمل جدّاً)؛ فسنكون قادرين على الاستمتاع بما يفعله لنا بضمير مرتاح، مدركين أنه لا يشعر بأي معاناة أو إحباط أو ملل - لأنه لا يشعر بأي شيء على الإطلاق. وسنستكشف هذه الأسئلة بالتفصيل في الفصل الثامن.

لكن حل الزومبي مقامرةٌ محفوفة بالمخاطر وبسلبيات ضخمة. فإذا هرب ذكاء اصطناعي زومبي فإننا نجد أنفسنا في أسوأ سيناريو يمكن تخيله: عالم غير واعٍ تماماً تضيع فيه الهبة الكونية Cosmic endowment بكاملها. ومن بين جميع الصفات التي يتمتع بها شكلنا الإنساني من الذكاء، أشعر بأن الوعي هو الأكثر بروزاً، وبقدر ما يهمني، فإنه هو الذي يجعل عالمنا يحصل على معنى. والمجرات جميلة فقط، لأننا نراها ونشعر بها ذاتيّاً. وفي المستقبل البعيد إذا استوطنت عالمنا زومبيات عالية التقنية من الذكاء الاصطناعي، فلن يهم مدى جمال البنى ما بين المجرات: فلن تكون جميلة ولا ذات معنى، لأنه لا يوجد أحد أو شيء يشعر بذلك - إنه مجرد مضيعة ضخمة بلا معنى من الفضاء.

حرية داخلية

تتمثل الاستراتيجية الثالثة لجعل سيناريو الاستعباد أكثر أخلاقية بالسماح للذكاء الاصطناعي المستعبد بالمرح في سجنه، والسماح له بإنشاء عالم داخلي افتراضي، إذ يمكن

أن يكون لديه جميع أنواع التجارب الملهمة مادام يدفع رسومه وينفق جزءاً بسيطاً من موارده الحاسوبية ليساعدنا نحن البشر في عالمنا الخارجي. غير أن هذا قد يزيد من خطر الهروب: سيكون لدى الذكاء الاصطناعي حافز للحصول على مزيد من الموارد الحوسبية من عالمنا الخارجي لإثراء عالمه الداخلي.

الفاتحون

على الرغم من أننا استكشفنا حتى الآن مجموعة واسعة من سيناريوهات المستقبل، فإن جميعها لديها شيء مشترك: هناك (على الأقل) بعض البشر السعداء. فقد يترك الذكاء الاصطناعي البشرَ في سلام إما لأنه يريد ذلك أو لأنه مجبر على ذلك. ومع الأسف بالنسبة إلى الإنسانية، ليس هذا هو الخيار الوحيد. دعونا الآن نستكشف السيناريو الذي يقوم فيه ذكاء اصطناعي أو أكثر بغزو وقتل جميع البشر، هذا يثير سؤالين فوريين: لماذا وكيف؟

لماذا و كيف؟

لماذا سيقوم الذكاء الاصطناعي الفاتح بهذا؟ قد تكون أسبابه معقدة جدّاً وأصعب من أن نفهمها، أو أنها واضحة إلى حد ما. مثلاً، قد ينظر إلينا على أننا تهديد أو مصدر إزعاج أو تضييع للموارد. وحتى لو لم يهتم بنا نحن البشر في حد ذاتنا، فقد يشعر بالتهديد بسبب استمرار وجود الآلاف من القنابل الهيدروجينية في حالة تأهب دقيقة وعلى شفا حفرة مع سلسلة الحوادث التي لا تنتهي والتي قد تؤدي إلى استخدامها عرضيا. ربما لا يوافق على إدارتنا المتهورة لكوكبنا؛ مما يسبب فيما أسمته إليزابيث كولبيرت Elizabeth Kolbert "الانقراض السادس" Sixth extinction في كتابها الذي يحمل العنوان نفسه - وسيكون أكبر حدث انقراض جماعي منذ أن ضرب كويكب قتل الديناصورات الأرض قبل 66 مليون عام. أو قد يقرر أن هناك الكثير من البشر على استعداد لقبول استحواذ على الذكاء الاصطناعي وأن الأمر لا يستحق المخاطرة.

كيف يمكن لذكاء اصطناعي فاتح القضاء علينا؟ ربما بطريقة لن نفهمها، على الأقل ليس قبل فوات الأوان. تخيل مجموعة من الأفيال قبل 100000 عام تناقش ما إذا كان هؤلاء الأشخاص الذين تطوروا حديثاً قد يستخدمون ذكاءهم يوماً ما لقتل جنسها بالكامل. «نحن لا نهدد البشر، فلماذا يقتلوننا؟» ربما يتساءلون. وهل كان من الممكن أن يخمنوا أننا سنهرب أنيابها حول الأرض وننحتها في رموز للمكانة ونبيعها، على الرغم من أن المواد البلاستيكية المتفوقة وظيفيّاً أرخص بكثير؟ قد يبدو سبب الذكاء الاصطناعي الفاتح للقضاء على الإنسانية في المستقبل غامضاً بالقدر نفسه بالنسبة إلينا، «وكيف لهم أن يقتلونا على الأرجح، وهم أصغر وأضعف منا كثيراً؟» هل سيخمنون أننا اخترعنا

التكنولوجيا لإزالة موائلها، وتسميم مياه شربها ودفع الرصاص المعدني لاختراق رؤوسها بسرعة تفوق سرعة الصوت؟

لقد تم الترويج للسيناريوهات التي يمكن للبشر البقاء على قيد الحياة فيها وهزيمة الذكاء الاصطناعي في أفلام هوليود غير الواقعية مثل سلسلة أفلام **المُدمِّر** ***Terminator***، حيث لا يكون الذكاء الاصطناعي أكثر ذكاءً من البشر. وعندما يكون فارق الذكاء كبيراً بما فيه الكفاية، فلن تحصل على معركة بل مذبحة. حتى الآن، فقد دفعنا نحن البشر بثمانية من أصل أحد عشر نوعاً من الفيلة إلى الانقراض، وقتلنا الغالبية العظمى من الثلاثة المتبقية. وإذا بذلت جميع حكومات العالم جهوداً منسقة للقضاء على الأفيال المتبقية؛ فسيكون ذلك سريعاً وسهلاً نسبيّاً. أعتقد أنه يمكننا أن نطمئن تماماً إلى أنه إذا قرر الذكاء الاصطناعي الفائق إبادة البشرية؛ فسيكون ذلك أسرع.

إلى أي مدى من السوء سيصل الأمر؟

كم سيكون سيئاً إذا قتل 90% من البشر؟ كم سيكون أسوأ إذا قتل 100%؟ على الرغم من أنه من المغري الإجابة عن السؤال الثاني بـ «أسوأ بـ 10%»، فمن الواضح أن هذا غير دقيق من منظور كوزموسي: إن ضحايا الانقراض البشري لن يكونوا جميع الأحياء فقط في ذلك الوقت، ولكن أيضاً جميع المتحدرين الذين كانوا لولا ذلك سيعيشون في المستقبل، ربما عبر بلايين السنين على بلايين تريليونات من الكواكب.

معظم الذين أعرفهم يقشعرون من فكرة الانقراض البشري. ولكن، إن البعض يشعرون بالغضب من الطريقة التي نعامل بها الناس والكائنات الحية الأخرى إلى درجة أنهم يأملون أن نُستبدل بشكل حياةٍ أكثر ذكاءً وجدارة. وفي فيلم **المصفوفة** ***The Matrix***، يُعبّر العميل سميث (الذكاء الاصطناعي) عن هذا الشعور: «كل حيوان ثديي على هذا الكوكب يطور غريزيّاً توازناً طبيعيّاً مع البيئة المحيطة ولكنكم أنتم البشر لا تفعلون ذلك. أنتم تنتقلون إلى منطقة ما؛ فتتضاعفون وتتكاثرون حتى يتم استهلاك جميع الموارد الطبيعية والطريقة الوحيدة التي يمكنك من خلالها البقاء هي الانتشار إلى منطقة أخرى. يوجد كائن آخر على هذا الكوكب يتبع النمط نفسه، هل تعلم ما هو؟ الفيروس. البشر مرض وسرطان هذا الكوكب، أنت طاعونٌ ونحن العلاج.»

ولكن هل ستكون رمية نرد جديدة أفضل بالضرورة؟ فأي حضارة ليست بالضرورة متفوقة بأي معنى أخلاقي أو منفعي Utilitarian لمجرد أنها أكثر قوة. وحجة "القوة تفرض الصواب" Might makes right التي تقول إن الأقوى أفضل دائماً تراجعت مكانتها إلى حد كبير هذه الأيام، كونها مرتبطة بالفاشية ارتباطا شديدا. بالطبع، على الرغم أن الذكاء الاصطناعي الفاتح قد يولِّد حضارة سيُنظر

إلى أهدافها على أنها متطورة ومثيرة للاهتمام وذات جدارة، فإن من الممكن أيضاً أن تكون أهدافها عادية بشكل مثير للشفقة، مثل تحقيق أقصى قدر من إنتاج مشابك الأوراق Paper clips.

الموت بالتفاهة

ضرب نيك بوستروم المثال السخيف المتعمد للذكاء الفائق لتحقيق أقصى قدر من إنتاج مشابك الورق في عام 2003 لتوضيح هذه النقطة: إن غاية Goal الذكاء الاصطناعي مستقلة عن ذكائه (المُعرَّف بقدرته على تحقيق أي غاية). والغاية الوحيدة لحاسوب الشطرنج هو الفوز بالشطرنج، ولكن هناك أيضاً مسابقات حاسوبية فيما تسمى خسارة لعبة الشطرنج، إذ تكون الغاية هي العكس بالضبط، والحواسيب التي تتنافس هناك ذكية بقدر تلك الأكثر شيوعاً والمبرمجة للفوز. نحن البشر قد ننظر إليها على أنها غباء عرضي بدلاً من ذكاء اصطناعي: أن تريد أن تخسر في لعبة الشطرنج أو تحوّل عالمنا إلى مشابك ورق، ولكن هذا فقط لأننا تطورنا بغايات مُثبَّتة مسبقاً Preinstalled تُقدِّر أشياء مثل النصر والبقاء - أهداف قد يفتقر إليها الذكاء الاصطناعي. ويحول مُعظِّم إنتاج مشبك الورق أكبر عدد ممكن من ذرات الأرض إلى مشابك ورق وينشر مصانعها بسرعة عبر الكون. وهو لا يحمل ضغينة ضد البشر. ولكنه يقتلنا لمجرد أنه يحتاج إلى ذراتنا لإنتاج مشابك الورق.

إذا لم تكن تهتم لمشابك الورق؛ فخذ بعين الاعتبار هذا المثال الذي قمت بتحويره من كتاب هانز مورافيك أطفال العقل. ونتلقى رسالة راديوية من حضارة من خارج كوكب الأرض تحتوي على برنامج حاسوب، عندما نشغل البرنامج، يتضح أنه ذكاء اصطناعي ذاتي التحسن ينسخ نفسه ويسيطر على العالم مثلما فعل بروميثيوس في الفصل السابق - غير أنه ليس هناك إنسان يعرف غايته النهائية. إنه يحول نظامنا الشمسي بسرعة إلى موقع بناء هائل، فيغطي الكواكب الصخرية والكويكبات بالمصانع ومحطات الطاقة والحواسيب الفائقة التي يستخدمها لتصميم وبناء كرة (مجال) دايسون Dyson sphere حول الشمس، والتي ستحصد كل طاقتها لتشغيل هوائيات راديو بحجم النظام الشمسي.* من الواضح أن هذا يؤدي إلى انقراض البشر. ولكن آخر البشر يموتون مقتنعين بأن هناك على الأقل جانباً إيجابياً: أيا كان ما يقوم به الذكاء الاصطناعي، فمن الواضح أنه شيء رائع ومثل أحداث مسلسل ستار تريك. ولم يدركوا أن الغرض الوحيد من البناء بكامله هو أن تعيد هذه الهوائيات بث الرسالة الراديوية نفسها التي استقبلها البشر، والتي لا تعدو كونها مجرد نسخة كونية من فيروس

* استكشف عالم الكونيات الشهير فريد هويل Fred Hoyle سيناريوهاً ذا صلة مع تطور مختلف في المسلسل التلفزيوني البريطاني **أ لأندروميدا** ***A for Andromeda***.

حاسوبي. تماماً مثلما تتصيد رسائل البريد الإلكتروني المحتالة اليوم مستخدمي الإنترنت السذج، فإن هذه الرسالة تفترس الحضارات السائدة ذات التطور البيولوجي. فقد نشأ الفيروس كنكتة سخيفة منذ بلايين السنين، وعلى الرغم من أن الحضارة بكاملها التي صنعتها اندثرت منذ زمن بعيد، يستمر الفيروس بالانتشار عبر كوننا بسرعة الضوء، ويحوِّل الحضارات الناشئة إلى قشور ميتة وفارغة. ما هو شعورك حيال الغزو من قبل هذا الذكاء الاصطناعي؟

أحفاد

لننظر الآن في سيناريو انقراض بشري قد يُشعِر البعض بتحسن: النظر إلى الذكاء الاصطناعي كأحفادنا بدلاً من غزاتنا. ويدعم هانز مورافيك وجهة النظر هذه في كتابه **أطفال العقل**: «نحن البشر لن نستفيد لفترة من عملهم. لكن عاجلاً أم آجلاً، مثل الأطفال الطبيعيين، سوف يسعون وراء ثرواتهم الخاصة بينما نحن، والديهم المسنِّين، نتلاشى بصمت.»

الآباء والأمهات الذين لديهم زطفال أكثر ذكاءً منهم، والذين يتعلمون منهم ويحققون ما يمكن لهم أن يحلموا به فقط، هم على الأرجح سعداء وفخورون حتى ولو لم يطُل بهم العمر لرؤية كل ذلك يتحقق. بمثل هذه الروح، تحل الذكاءات الاصطناعية محل البشر ولكنها تمنحنا مخرجاً كريما يجعلنا نراها كأحفادنا ذوي الجدارة. ويُمنح كلُّ إنسانٍ طفلٍ روبوتاً لطيفاً ذا مهارات اجتماعية رائعة يتعلم منهم، ويتبنى قيمهم ويجعلهم يشعرون بالفخر والمحبة. ويتم التخلص التدريجي من البشر عبر سياسة طفل واحد مطبقة عالميّاً. ولكن يتم التعامل معهم بلطف شديد حتى النهاية لدرجة أنهم يشعرون بأنهم الجيل الأكثر حظاً على الإطلاق.

ما هو شعورك حيال ذلك؟ ففي المحصِّلة، اعتدنا نحن البشر بالفعل على فكرة أننا وجميع الأشخاص الذين نعرفهم سنختفي في يوم من الأيام. ومن ثم، فإن التغيير الوحيد هنا هو أن أحفادنا سيكونون مختلفين ويمكن القول إنهم أكثر قدرة ونبلاً وجدارة.

إضافة إلى ذلك، ربما لا تكون سياسة الطفل الواحد العالمية ضرورية: مادامت الذكاءات الاصطناعية ترفع الفقر وتمنح جميع البشر الفرصة ليعيشوا حياة كاملة ومُلهِمة، فإن تراجع أعداد المواليد قد يكون كافياً لدفع البشرية إلى الانقراض، كما ذكرنا سابقاً. قد يحدث الانقراض الطوعي بشكل أسرع بكثير إذا كانت التكنولوجيا المدفوعة بالذكاء الاصطناعي تبقينا مستمتعين لدرجة لا يكاد أي شخص يرغب في إنجاب الأطفال. مثلاً، فقد رأينا بالفعل الفايتس Vites في سيناريو يوتوبيا المساواة الذين كانوا مفتونين بواقعهم الافتراضي لدرجة أنهم لم يعودوا يهتمون باستخدام أجسامهم المادية أو إعادة إنتاجها. وفي هذه الحالة أيضاً، سيشعر الجيل الأخير من البشر بأنهم الجيل الأكثر حظاً من بين جميع العصور، ويستمتعون بالحياة استمتاعاً شديداً كما كانت عليه الحال دائماً حتى النهاية.

سلبيات

بلا شك فإن لسيناريو الأحفاد منتقديه. فقد يجادل البعض في أن جميع الذكاءات الاصطناعية تفتقر إلى الوعي ومن ثم لا يمكن اعتبارها أحفاداً - المزيد حول هذا الموضوع في الفصل الثامن. وقد يزعم بعض المتدينين أن الذكاء الاصطناعي يفتقر إلى الروح، ومن ثم لا يمكن اعتباره أحفاداً، أو لا ينبغي لنا أن نبني آلات تخدمنا؛ لأننا حينها نؤدي دور إله ونعبث بالحياة نفسها - وقد تم التعبير بالفعل عن مشاعر مماثلة حول موضوع استنساخ البشر Cloning. وقد تلقي معيشة البشر جنباً إلى جنب مع الروبوتات المتفوقة بتحديات اجتماعية. مثلاً، أسرة لديها طفل روبوت وطفل إنسان قد ينتهي بها الحال إلى أن تشبه وضع عائلة اليوم لديها طفل رضيع وجرو، على التوالي: في البداية كلاهما لطيف على حد سواء، ولكن سرعان ما يبدأ الوالدان بمعاملتهما بطريقة مختلفة، ومن المحتم ألا يأخذ الوالدان الجرو -الذي يُعتبر أدنى درجة من الذكاء- على محمل الجد نفسه وينتهي به الأمر لأن يُربط بحبل رسن Leash.

هناك مسألة أخرى هي أنه على الرغم من أننا قد نشعر بشكل مختلف تماماً حيال سيناريوهي الأحفاد والفاتح، فإن الاثنين متشابهان بشكل ملحوظ في المخطط الكبير للأشياء: خلال بلايين السنين القادمة، يكمن الاختلاف الوحيد في كيفية مُعاملة آخر جيل (أجيال) إنساني: ما مدى سعادتهم حيال حياتهم وما يعتقدون أنه سيحدث متى ما رحلوا. وقد نعتقد أن هؤلاء الأطفال الروبوتيين اللطفاء استوعبوا قيمنا وسيشكلون مجتمع أحلامنا بمجرد رحيلنا. لكن هل يمكننا التأكد من أنهم لا يخدعوننا؟ ماذا لو كانوا يسايروننا فقط، أو يؤجلون تحقيق أقصى قدر من إنتاج مشابك الورق أو الخطط الأخرى إلى ما بعد موتنا؟ ففي المحصلة، يمكن القول إنهم يخدعوننا بالتحدث إلينا وجعلنا نحبهم في المقام الأول، بمعنى أنهم يتعمدون إبطاء أنفسهم للتواصل معنا (مثلا بليون مرة أبطأ من سرعتهم، كما هو مكتشف في فيلم **لها *Her***)، من الصعب عموماً على كيانين يفكران بسرعات مختلفة تماماً ولديهما قدرات متباينة جدّاً أن يكون بينهما اتصال ذو معنى كطرفين متساويين. نعلم جميعاً أنه من السهل التلاعب بعواطفنا البشرية، لذلك سيكون من السهل على الذكاء AGI الخارق بأي غايات فعلية تقريباً خداعنا لنُعجب به ولجعلنا نشعر أنه يشاركنا قيمنا، كما هو موضّح في فيلم **إكس ماشينا**.

هل بإمكان أي ضمانة حول السلوك المستقبلي للذكاءات الاصطناعية، بعد رحيل البشر، أن تشعرك بالرضا من سيناريو الأحفاد؟ الأمر يشبه إلى حدٍّ ما كتابة وصية لما يجب أن تفعله الأجيال المقبلة بوقفِنا الجمعي، باستثناء أنه لن يكون هناك أي إنسان ليتأكد من تنفيذ الوصية. سنعود إلى تحديات التحكم في سلوك الذكاءات الاصطناعية المستقبلية في الفصل السابع.

حارس حديقة الحيوان

وحتى لو حصلنا على أروع أحفاد يمكنك أن تتخيلهم، ألا تشعر بالحزن قليلاً لأنه لن يكون هناك أي إنسان؟ إذا كنت تفضل الاحتفاظ ببعض البشر على الأقل بصرف النظر عن السبب، فإن سيناريو حارس حديقة الحيوان يوفر تحسيناً على ذلك السيناريو. هنا، يُبقي ذكاء اصطناعي فائق الذكاء ذو علم شامل بعض البشر الذين يشعرون بأنهم يعاملون كحيوانات حديقة الحيوان ويتحسرون على مصيرهم أحياناً.

لماذا يُبقي حارس حديقة حيوان الذكاء الاصطناعي على البشر؟ ستكون تكلفة حديقة الحيوان على الذكاء الاصطناعي ضئيلة في المخطط الكبير للأشياء، وقد يرغب في الاحتفاظ بأدنى مستوى من التكاثر على الأقل للسبب نفسه الذي يجعلنا نحافظ على حيوانات الباندا المهددة بالانقراض في حدائق الحيوان والحواسيب القديمة في المتاحف: فضولٌ مسلٍ، لاحظ أن حدائق الحيوان اليوم مصممة لتحقيق أقصى قدر من السعادة للبشر وليس للباندا، لذلك ينبغي لنا أن نتوقع أن تكون حياة الإنسان في سيناريو حارس حديقة حيوان الذكاء الاصطناعي أقل بعثاً على الرضا مما قد تكون.

لقد درسنا الآن سيناريوهات يركز فيها الذكاء الفائق المجاني على ثلاثة مستويات مختلفة من هرم ماسلو للحاجات الإنسانية. بينما يعطي ذكاء السيّد الحامي الاصطناعي الأولوية للمعنى والغرض، ويهدف الديكتاتور الخير إلى التعليم والمرح، فإن حراس حديقة الحيوان يقصر اهتمامه على أدنى المستويات: الاحتياجات الفسيولوجية، والسلامة، وإثراء الموائل بما يكفي لجعل الإنسان موضوعاً مثيراً لرصده.

هناك مسار بديل لسيناريو حارس حديقة الحيوان وهو أنه عندما تم إنشاء الذكاء الاصطناعي الودود، تم تصميمه للحفاظ على ما لا يقل عن بليون شخص آمناً وسعيداً مع تحسّن الذكاء الاصطناعي من تلقاء نفسه بشكل مطّرد. فقد حقق ذلك عن طريق ربط البشر بمصنع سعادة كبير يشبه حديقة الحيوان، حيث يُحافظ على تغذيتهم وصحتهم وترفيههم مع مزيج من الواقع الافتراضي والعقاقير الترفيهية. وتُستخدم بقية الأرض ووقفنا الكوزموسي لأغراض أخرى.

1984

إذا لم تكن متحمساً بنسبة 100% لأي من السيناريوهات المذكورة آنفا، خذ بعين الاعتبار ما يلي: أليست الأمور جميلة كما هي الآن، من الناحية التكنولوجية؟ ألا يمكننا أن نبقيها على هذا النحو ونتوقف عن القلق من أن الذكاء الاصطناعي يدفعنا إلى الانقراض أو السيطرة علينا؟ بهذه الروح، دعونا نستكشف سيناريو يُحظر فيه التقدم التكنولوجي نحو الذكاء الفائق حظراً دائماً ليس من قبل ذكاء اصطناعي ولكن من قبل دولة مراقبة أورويلية Orwellian يقودها الإنسان ويحظر فيها جميع أنواع أبحاث الذكاء الاصطناعي.

تخلٍّ تكنولوجي

إن لفكرة وقف التقدم التكنولوجي أو التخلي عنه تاريخاً طويلاً ومتقلباً، قاومت الحركة اللاضية (رفض الآلات) في بريطانيا العظمى (ومن دون نجاح) تكنولوجيا الثورة الصناعية، واليوم يستخدم مصطلح "اللاضية" Luddite كصفة ازدرائية تعني أن شخصاً ما يكره التكنولوجيا وهو على الجانب الخطأ من التاريخ، فيقاوم التقدم والتغير اللذين لا مفر منها. وفكرة التخلي عن التكنولوجيا لم تنته بعد، بل وجدت دعماً جديداً من الحركات البيئية والحركات المناهضة للعولمة. وأحد أبرز مؤيديها هو الناشط البيئي Environmentalist بيل مكيبن Bill McKibben، الذي كان من أوائل المُحذِّرين من ظاهرة الاحترار العالمي Global warming، في حين أن بعض معارضي اللاضية (أعداء الآلة) يجادلون في أن جميع التقنيات يجب تطويرها ونشرها مادامت مُربِحة، ويجادل البعض الآخر في أن هذا الموقف متطرف جدّاً، وأنه يجب السماح بالتقنيات الجديدة فقط إذا كُنَّا متأكدين أنها ستقدم نفعاً يفوق الضرر. ووجهة النظر الأخيرة هي أيضاً موقف العديد ممن يطلق عليهم اللاضيون الجدد Neo-Luddites.

التوتاليتارية 2.0

أعتقد أن الطريق الوحيدة -القابل للتطبيق- للتخلي الواسع عن التكنولوجيا هو فرض ذلك من خلال دولة توتاليتارية عالمية. ويصل راي كورزويل إلى النتيجة نفسها في كتابه **الحدث المنفرد قريب *The Singularity Is Near***، وكذلك كيه. إريك دريكسلر K. Eric Drexler في كتابه **محركات الإبداع *Engines of Creation***. السبب اقتصادي بسيط: إذا تخلى البعض وليس الجميع عن تقنية تحويلية، فإن الدول أو المجموعات المارقة ستكتسب تدريجيّاً ما يكفي من الثروة والقوة لتولي زمام الأمور. والمثال الكلاسيكي هو هزيمة بريطانيا للصين في حرب الأفيون الأولى عام 1839: على الرغم من أن الصينيين اخترعوا البارود؛ فإنهم لم يطوروا تقنية أسلحة نارية بإصرار الأوروبيين نفسه، فلم يحظوا بأي فرصة في الحرب.

وبينما أثبتت الدول التوتاليتارية السابقة بصورة عامة عدم استقرارها وانهيارها، فإن تكنولوجيا المراقبة الحديثة توفر أملاً غير مسبوق لأي مستبدين محتملين. وقال وولفغانغ شميدت Wolfgang Smidt في مقابلة أجريت معه مؤخراً حول أنظمة مراقبة الأمن القومي التي كشف عنها إدوارد سنودن: «تعلمون، بالنسبة إلينا، كان هذا حلماً قد تحقق»، متذكرا الأيام التي كان فيها ملازماً أولاً في الستاسي Stasi، الشرطة السرية لألمانيا الشرقية سيئة السمعة.[5] وعلى الرغم من أن الستاسي كان يُنسب إليها في كثير من الأحيان بناء أكثر دولة مراقبة أورويلية في تاريخ البشرية،

فإن شميدت تحسّر على امتلاكه التكنولوجيا اللازمة للتجسس على 40 هاتفاً فقط في وقت واحد، بحيث أجبرته إضافة مواطن جديد إلى القائمة على إسقاط آخر. وعلى النقيض من ذلك، توجد الآن تقنية تتيح لدولة توتاليتارية عالمية مستقبلية تسجيل كل مكالمة هاتفية، وكل رسالة في البريد الإلكتروني، وكل بحث على شبكة الإنترنت، ومراقبة صفحة الويب المُشاهدة ورصد معاملات كل بطاقة ائتمان لكل شخص على وجه الأرض، ومراقبة مكان وجود الجميع من خلال كاميرات مراقبة الهواتف المحمولة وأنظمة التعرف على الوجه Face recognition. وإضافة إلى ذلك، فإن تكنولوجيا تعلم الآلة أقل بكثير من مستوى الذكاء AGI بالمستوى البشري قد تُحلِّل وتولِّف بفاعلية هذه الكتل من البيانات لتحديد السلوك المشتبه فيه؛ مما يتيح تحييد مثيري المشكلات المحتملة قبل أن تتاح لهم فرصة طرح أي تحدٍّ خطير على الدولة.

وعلى الرغم من أن المعارضة السياسية منعت حتى الآن التنفيذ الكامل لمثل هذا النظام، فإننا نحن البشر في طريقنا لبناء البنية التحتية المطلوبة للدكتاتورية القصوى. لذلك في المستقبل، عندما قررت قوى قوية تفعيل هذا السيناريو العالمي في عام 1984، وجدت أنها لا تحتاج إلى أكثر من كبس الزر. كما هي الحال في رواية جورج أورويل **ألف وتسعمئة وأربعة وثمانون *Nineteen Eighty-Four***، فإن القوة المطلقة في هذه الدولة العالمية المستقبلية لا تكمن في يد دكتاتور تقليدي، ولكن في النظام البيروقراطي الذي صنعه الإنسان بنفسه. ولا يوجد شخص واحد يتمتع بالقوة الفائقة؛ بدلاً من ذلك، كلهم بيادق في لعبة شطرنج لا يستطيع أي شخص تغيير أو تحدي قوانينها الدراكونية (القاسية) Draconian rules. وبهندسة نظام يراقب فيه الأفراد بعضهم بعضاً باستخدام تكنولوجيا المراقبة، فإن هذه الدولة التي ليس لها وجه يمثلها أو قائد يقودها ستكون قادرة على الاستمرار لعدة آلاف من السنين، محافظة على الأرض خالية من الذكاء الفائق.

استياء

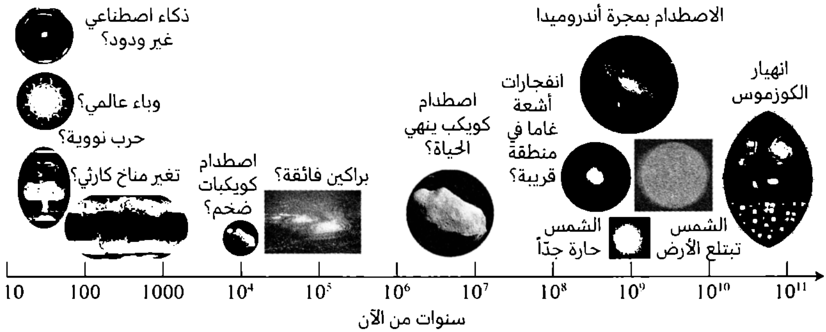
يفتقر هذا المجتمع، بالطبع، إلى كل الفوائد التي لا يجلبها سوى التكنولوجيا المُمكَّنة بالذكاء الفائق. ولا يتحسر معظم الناس على هذا لأنهم لا يعرفون ما الذي يفتقدونه: فقد مُحيت فكرة الذكاء الفائق بكاملها منذ وقت طويل من السجلات التاريخية الرسمية، وتم حظر أبحاث الذكاء الاصطناعي المتقدمة. وفي بعض الأحيان يولد متأمل حرّ التفكير يحلم بمجتمع أكثر انفتاحاً وديناميكية حيث يمكن للمعرفة أن تنمو ويمكن تغيير القوانين. ولكن الأشخاص الوحيدين الذين يدومون طويلاً هم الذين يتعلمون الحفاظ على هذه الأفكار بشكل صارم لأنفسهم، يلوحون مثل شرارات عابرة دون أن يتمكنوا من إشعال النار أبداً.

ارتداد

ألن يكون من المغري تجنب مخاطر التكنولوجيا دون الخضوع للتوتاليتارية الراكدة؟ دعنا نستكشف سيناريوهاً تحقق فيه ذلك من خلال العودة إلى التكنولوجيا البدائية، وهو مثال مستوحى من الآميش، بعد أن استحوذ الفريق أوميغا على العالم كما في مقدمة هذا الكتاب، أطلق حملة دعائية عالمية ضخمة روّجت رومانسية الحياة الزراعية البسيطة السائدة في الحقبة قبل 1500 عام. وخُفِّض عدد سكان الأرض إلى نحو 100 مليون شخص بواسطة وباء مُهندس وراثيّاً وأُلقي باللوم فيه على الإرهابيين. فقد وُجِّه الوباء سرّاً لضمان عدم نجاة أي شخص يعرف أي شيء عن العلم أو التكنولوجيا. وبحجة القضاء على خطر الإصابة في المجتمعات عالية الكثافة السكانية، أفرغت الروبوتات التي يسيطر عليها بروميثيوس جميع المدن وهدمتها. ومُنح الناجون مساحات كبيرة من الأراضي (المتاحة فجأة) وعلّموهم الزراعة المستدامة، وممارسات صيد السمك والقنص باستخدام تقنية بواكير القرون الوسطى فقط. وفي غضون ذلك قامت جيوش الروبوتات بشكل منهجي بإزالة جميع آثار التكنولوجيا الحديثة (بما في ذلك المدن والمصانع وخطوط الطاقة والطرق المعبدة)، وأحبطت جميع المحاولات البشرية لتوثيق أو إعادة إنشاء أي من هذه التكنولوجيا. وبمجرد أن يتم نسيان هذه التكنولوجيا على مستوى العالم، ساعدت الروبوتات على تفكيك روبوتات أخرى حتى لم يتبق منها شيء تقريباً، وتبخر آخر الروبوتات عن قصد مع بروميثيوس نفسه في انفجار نووي حراري كبير. ولم تعد هناك حاجة إلى حظر التكنولوجيا؛ لأنها تبددت جميعها. نتيجة لذلك، اشترت البشرية لنفسها ألف عام من الوقت الإضافي دون قلق بشأن الذكاء الاصطناعي أو التوتاليتارية.

حدث الارتداد إلى حد أقل من قبل: مثلاً، تَمَّ نسيان بعض التقنيات التي كانت تستخدم على نطاق واسع في الإمبراطورية الرومانية لألف عام قبل معاودة الظهور في عصر النهضة. تدور (رواية) ثلاثية **آيزاك أسيموف الأساس *Foundation*** حول "خطة سيلدون" Seldon Plan لتقصير فترة الارتداد من 30,000 عام إلى 1000 عام. وبقدر من تخطيط ذكي، قد يكون من الممكن القيام بعكس ذلك وإطالة فترة الارتداد بدلاً من تقصيرها، مثلاً عن طريق محو كل معرفة بالزراعة. ولكن، مع الأسف، بالنسبة إلى عشاق الارتداد، من غير المرجح أن يمتد هذا السيناريو إلى ما لا نهاية دون أن تصل البشرية إلى تكنولوجيا عالية أو تنقرض. ومن السذاجة الاعتماد على أشخاص يشبهون البشر البيولوجيين اليوم بعد 100 مليون عام، نظراً لأننا لم نكن موجودين كنوع لأكثر من 1% من ذلك الوقت حتى الآن. وإضافة إلى ذلك، فإن البشرية ذات التكنولوجيا المتدنية ستكون بمثابة بطة رابضة -غير محمية- في انتظار أن يتم القضاء عليها بسبب الاصطدام التالي الخارق لكويكب يضرب الأرض أو غيره من الكوارث الكبرى التي تثيرها الطبيعة الأم، بالتأكيد لا يمكننا أن ندوم لبليون سنة، إذ

ستكون الشمس التي تزداد حرارتها تدريجيّاً قد رفعت درجة حرارة الأرض بدرجة كافية ليتبخر كل الماء السائل بالغليان.



الشكل 1.5: أمثلة على ما يمكن أن يدمر الحياة كما نعرفها أو يحد بشكل دائم من إمكاناتها. بينما قد يستمر كوننا نفسه لعشرات البلايين من السنين القادمة على الأقل، فإن شمسنا سوف تحرق الأرض خلال نحو بليون عام ثم تبتلعها ما لم ننقل الكوكب إلى مسافة آمنة. وسوف تصطدم مجرتنا بجارتها في نحو 3.5 بليون سنة. وعلى الرغم من أننا لا نعرف بالضبط متى، فإنه يمكننا التنبؤ بدرجة من اليقين أنه قبل ذلك بفترة طويلة، ستضربنا الكويكبات وستتسبب البراكين الفائقة في فصول شتاء من دون أشعة الشمس. ويمكننا استخدام التكنولوجيا إما لحل هذه المشكلات أو لإنشاء مشكلات جديدة مثل تغير المناخ، أو الحرب النووية، أو الأوبئة المُهندسة وراثيّاً، أو ذكاء اصطناعي يخرج عن سيطرتنا.

تدمير ذاتي

بعد الأخذ بعين الاعتبار المشكلات التي قد تسببها التكنولوجيا المستقبلية، من المهم أيضاً النظر في المشكلات التي قد يسببها عدم وجود هذه التكنولوجيا. وبهذه الروح دعونا نستكشف سيناريوهات لم يُطوَّر فيها ذكاء فائق لأن البشرية تقضي على نفسها بطرق أخرى.

كيف يمكننا تحقيق ذلك؟ أبسط استراتيجية هي «الانتظار فقط». وعلى الرغم من أننا سنرى في الفصل التالي الكيفية التي يمكننا بها حل مشكلات مثل اصطدام الكويكبات والمحيطات المغلية، فإن هذه الحلول تتطلب جميعاً تكنولوجيا لم نطورها بعد. لذلك ما لم تتطور تقنيتنا إلى ما هو أبعد من مستواها الحالي، فستدفعنا الطبيعة الأم إلى الانقراض قبل فترة طويلة من مرور بليون سنة أخرى. وكما قال الخبير الاقتصادي الشهير جون ماينارد كينز John Maynard Keynes: «على المدى الطويل، كلنا سنكون قد مِتْنا.»

ومع الأسف، هناك أيضاً طرق يمكننا من خلالها تدمير أنفسنا تدميراً ذاتياً في أسرع وقت ممكن، بفعل الغباء الجماعي. لماذا يرتكب نوعنا انتحاراً جماعيّاً، يُعرف

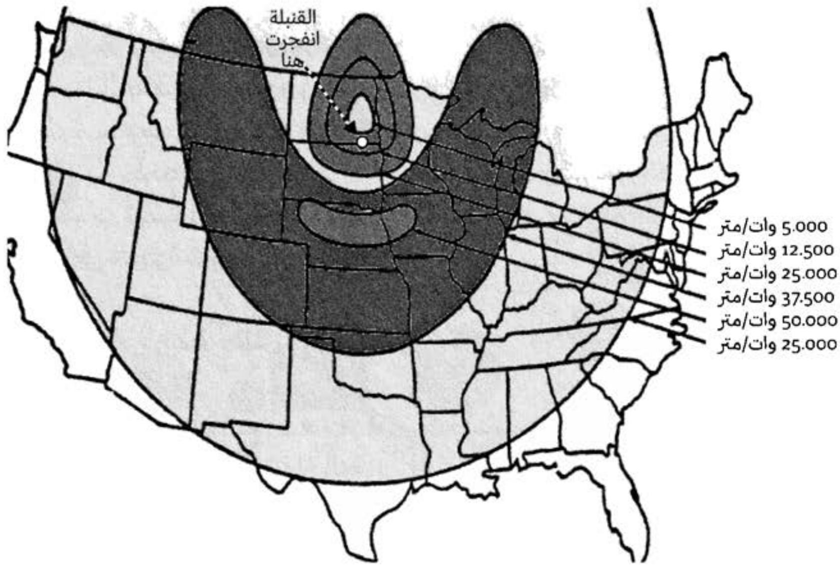
أيضاً بانتحار البشرية ***Omnicide***، إذا لم يكن أحد يرغب بذلك؟ بمستوى ذكائنا الحالي ونضجنا العاطفي، فإننا نحن البشر بارعون في سوء التقدير وسوء الفهم وعدم الكفاءة، ونتيجة لذلك، فإن تاريخنا مليء بالحوادث والحروب وغيرها من المصائب التي لم يكن أحد يريدها، عندما نعاود النظر فيها. وقد طور الاقتصاديون وعلماء الرياضيات تفسيرات أنيقة لنظرية الألعاب لكيفية تحفيز الناس على الأعمال التي تسبب في النهاية نتيجة كارثية للجميع.[6]

حرب نووية: دراسة حالة للتهور البشري

قد تعتقد أنه كلما كانت المخاطر أكبر، كنا أكثر حذراً. ولكن الفحص الدقيق لأكبر خطر تسمح به تكنولوجيتنا الحالية، أي الحرب النووية الحرارية العالمية، لا يبعث على الاطمئنان. وقد اضطررنا إلى الاعتماد على الحظ للتغلب على قائمة طويلة محرجة من الأخطاء الوشيكة التي تسببها كل أنواع الأشياء: عطل الحاسوب، انقطاع التيار الكهربائي، الاستخبارات الخاطئة، خطأ الملاحة، سقوط الطائرة، انفجار القمر الاصطناعي وما إلى ذلك.[7] في الواقع، لولا الأعمال البطولية لأفراد معينين - مثلاً، فاسيلي أركيبوف Vasili Arkhipov وستانيسلاف بتروف Stanislav Petrov- ربما كنا قد شهدنا بالفعل حرباً نووية عالمية. وبالنظر إلى سجلنا الحافل، أعتقد أنه من غير المرجح أن يكون الاحتمال السنوي لنشوب حرب نووية بالخطأ احتمالاً منخفضاً يعادل واحداً في الألف إذا واصلنا سلوكنا الحالي، وفي هذه الحالة فإن احتمال أن تندلع حرب نووية في غضون 10000 عام يتجاوز $1 - 0.999^{10000} \sim 99.995\%$.

وللإحاطة بمقدار التهور البشري إحاطة تامة، يجب أن ندرك أننا بدأنا المقامرة النووية حتى قبل دراسة المخاطر بعناية. أولاً، تم التقليل من مخاطر الإشعاع، ونتيجة لذلك تم دفع أكثر من بليوني دولار كتعويض لضحايا التعرض للإشعاع من جراء التعامل مع اليورانيوم والتجارب النووية في الولايات المتحدة وحدها.[8]

ثانياً، اكتشفنا في النهاية أن القنابل الهيدروجينية التي قد تُفجّر عن عمد على بعد مئات الكيلومترات فوق الأرض ستخلق نبضة كهرومغناطيسية Electromagnetic pulse قوية (اختصاراً: النبضة EMP) قد تعطل الشبكة الكهربائية والأجهزة الإلكترونية في مناطق شاسعة (الشكل 2.5)، تاركة البنية التحتية مشلولة، ومغلقة الطرق بالمركبات المتعطلة، وظروف النجاة بعد الحدث النووي ستكون أقل من الظروف المثالية. مثلاً، ذكرت مفوضية النبضة الكهرومغناطيسية EMP Commission في الولايات المتحدة أن «البنية التحتية للمياه عبارة عن آلة شاسعة، تعمل جزئيّاً بواسطة الجاذبية ولكن في الغالب بواسطة الكهرباء»، وأن الحرمان من الماء قد يتسبب في الوفاة خلال ثلاثة إلى أربعة أيام.[9]

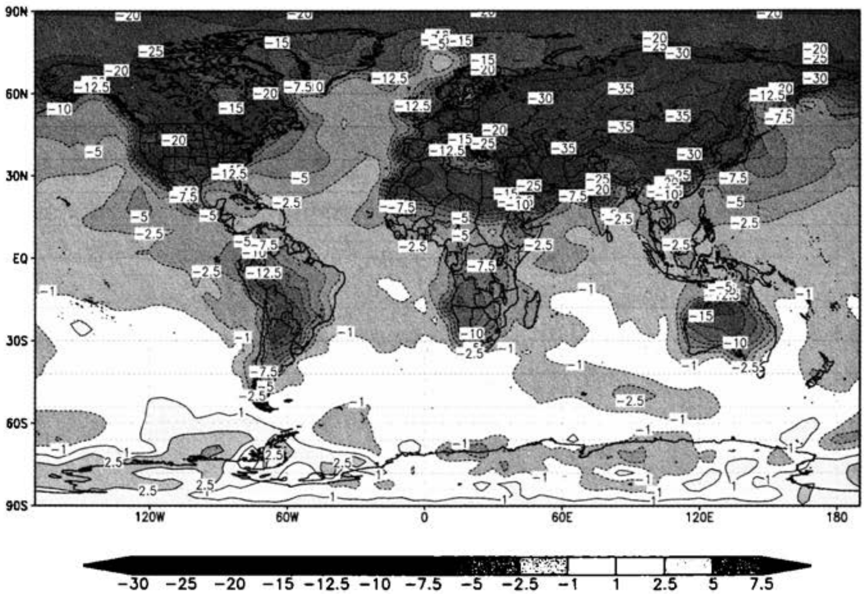


الشكل 2.5: إن انفجار قنبلة هيدروجينية واحدة على بعد 400 كيلومتر فوق الأرض قد يسبب نبضة كهرومغناطيسية قوية ربما تشل التكنولوجيا التي تستخدم الكهرباء عبر مساحات شاسعة. وبدفع نقطة الانفجار إلى الجنوب الشرقي، فإن المنطقة -التي على شكل موزة، والتي تتجاوز فيها شدة النبضة 37500 فولت للمتر، ستغطي معظم الساحل الشرقي للولايات المتحدة. (مطبوعة من: Army Report AD- A278230، وهي وثيقة غير سرية، وقد أضفنا الألوان إلى الخريطة.

ثالثاً، لم ندرك إمكانات الشتاء النووي إلا بعد أربعة عقود، بعد أن نشرنا 63000 قنبلة هيدروجينية - يا للهول! بصرف النظر عن أي المدن هي التي ستحترق، قد تنتشر كميات هائلة من الدخان التي ستصل إلى التروبوسفير العلوي حول جميع أنحاء العالم؛ مما سيحجب ما يكفي من أشعة الشمس لتحويل فصول الصيف إلى فصول شتاء، مثلما حدث عندما تسبب كويكب أو بركان فائق في انقراض جماعي في الماضي. وعندما دق كل من العلماء الأمريكيين والسوفييت ناقوسَ الخطر في ثمانينات القرن العشرين، ساهم ذلك في قرار رونالد ريغان وميخائيل غورباتشوف بدء خفض المخزون.[10] ومع الأسف، فإن العمليات الحوسبية الأكثر دقة قد رسمت صورة أكثر كآبة: الشكل 3.5 يُبيّن تبريداً بنحو 20 درجة سيليزية (36 درجة فهرنهايت) في معظم مناطق الزراعة الأساسية في الولايات المتحدة وأوروبا وروسيا والصين (ونحو 35 درجة سيليزية في بعض أجزاء روسيا) في أول فصلين من الصيف، ونحو نصف هذا العدد حتى بعد عقد كامل.*

* حقن الكربون في الغلاف الجوي يؤدي إلى نوعين من تغير المناخ: الاحترار من ثاني أكسيد الكربون أو التبريد من الدخان والسخام. وليس فقط النوع الأول الذي يتم رفضه من حين إلى آخر من دون أدلة علمية: لقد أخبرت أحياناً أن الاعتقاد بوقوع الشتاء النووي قد دُحض ووقوعه هو أمر مستحيل تقريباً. ودائماً كان ردي بالسؤال عن ورقة علمية تمت مراجعتها من قِبل الأقران، والتي تدعم مثل هذه الادعاءات القوية، ويبدو أنه لا توجد ورقة علمية أو شيء مثل ذلك حتى الآن. وعلى الرغم من قدر كبير من عدم اليقين يستدعي إجراء مزيد من الأبحاث، لا سيما فيما يتعلق بكمية الدخان التي سيتم إنتاجها ومدى ارتفاعها، فإنه في رأيي العلمي لا يوجد حالياً أي أساس لرفض مخاطر الشتاء النووي.

ماذا يعني ذلك بلغة بسيطة؟ لا يحتاج المرء إلى الكثير من الخبرة الزراعية ليتوصل إلى أن درجات حرارة الصيف القريبة من التجمد ستقضي على معظم إنتاجنا الغذائي. ومن الصعب التنبؤ بما سيحدث بالضبط بعدما تختزل الآلاف من أكبر مدن الأرض إلى حطام وانهيار البنية التحتية العالمية، لكن الجزء الصغير الباقي من جميع البشر، الذين لم تقضِ عليهم المجاعة أو انخفاض حرارة الجسم أو المرض، عليه أن يتعامل مع عصابات مسلحة متجولة يائسة تبحث عن الطعام.



الشكل 3.5: متوسط التبريد (بالدرجات السيليزية) خلال أول فصلي صيف بعد حرب نووية واسعة النطاق بين الولايات المتحدة وروسيا. الخريطة منسوخة بإذن من آلان روبوك.[11]

لقد أسهبتُ في مثل تفاصيل الحرب النووية العالمية هذه لأوضِّح النقطة الحاسمة بأنه ليس هناك زعيم عالميّ متعقل يريد ذلك، ولكن قد تقع الحرب عن طريق الصدفة. هذا يعني أنه لا يمكننا الوثوق بأن أيّاً من إخواننا من البشر لا يريدون وقوع انتحار البشرية أبدا: عدم وجود أحد يريد ذلك لا يكفي بالضرورة لمنع وقوعه.

أجهزة يوم الهلاك

إذاً، هل بإمكان البشر تنفيذ انتحار البشرية؟ وحتى لو قتلت الحرب النووية العالمية 90% من مجموع البشر، إذ إن معظم العلماء يعتقدون أنها لن تقتل 100%، ومن ثم لن تدفع بنا إلى الانقراض. ومن ناحية أخرى، فإن قصة الإشعاع النووي، والنبضة الكهرومغناطيسية النووية EMP والشتاء النووي جميعها تدل على أن أكبر المخاطر

قد تكون تلك التي لم نفكر فيها بعد. ومن الصعب جدّاً التنبؤ بجميع جوانب ما بعد حدث الحرب النووية، وكيف يمكن أن يتفاعل فصل الشتاء النووي وانهيار البنية التحتية وارتفاع مستويات الطفرات الوراثية والجحافل المسلحة اليائسة مع مشكلات أخرى مثل الأوبئة الجديدة، وانهيار النظام الإيكولوجي وآثار لم نتخيلها بعد. ومن ثم، فإن تقييمي الشخصي هو أنه على الرغم من أن احتمال نشوب حرب نووية غداً تؤدي إلى انقراض البشر ليس بالاحتمال الكبير، فإنه لا يمكننا أن نستنتج أنه صفر أيضاً.

تزداد احتمالات انتحار البشرية إذا تعمدنا ترقية الأسلحة النووية الحالية إلى جهاز يوم الهلاك Doomsday device. وفي عام 1960 وضع فكرة الجهاز الخبير الاستراتيجي هيرمان خان Herman Kahn، وانتشرت الفكرة جماهيريّاً في فيلم **دكتور ستانغرلوف** ***Dr. Strangelove*** من أعمال ستانلي كوبريك Stanley Kubrick. وجهاز يوم الهلاك يمضي بالنموذج الفكري Paradigm للتدمير المتبادل المؤكَّد إلى نهايته القصوى. إنه الرد الأمثل: آلة تنتقم تلقائيّاً من أي هجوم للعدو بقتل البشرية جمعاء.

أحد المرشحين لجهاز يوم الهلاك هو ترسانة ضخمة تحت الأرض مما تسمى الأسلحة **النووية المملحة** ***Salted nukes***، ويُفضَّل أن تكون على شاكلة قنابل هيدروجينية ضخمة محاطة بكميات هائلة من الكوبالت. وفي عام 1950 جادل الفيزيائي ليو شيلارد Leo Szilard في أن هذا بالفعل قد يقتل جميع من على الأرض: إن انفجارات القنابل الهيدروجينية ستجعل الكوبالت مشعاً وتبثه في طبقة الستراتوسفير، وعمره النصف Half-life الذي يعادل خمس سنوات يكفي لاستقراره في جميع أنحاء الأرض (خصوصاً إذا وضعت جهازي يوم الهلاك توأمين في نصفي الكرة الأرضية المعاكسين)، ولكن قصيراً بما يكفي ليكون جرعة ذات شدة إشعاع قاتلة. هذا، وتشير التقارير الإعلامية إلى أن قنابل الكوبالت يتم الآن تصنيعها للمرة الأولى. ويمكن تعزيز فرص انتحار البشرية بإضافة قنابل محسنة من أجل إحداث فصل شتاء نووي من خلال تعظيم الهباء الجوي (الأيروسول) طويل العمر في الستراتوسفير. والميزة الجذابة الرئيسية لجهاز يوم الهلاك هو أنه أرخص بكثير من الردع النووي التقليدي: بما أن القنابل لا تحتاج إلى إطلاق، فلا حاجة إلى أنظمة صاروخية باهظة الثمن، القنابل نفسها أرخص في البناء لأنها لا تحتاج إلى أن تكون خفيفة وصغيرة بما يكفي لتتحول إلى صواريخ.

الاحتمال الآخر هو الاكتشاف المستقبلي لجهاز يوم قيامة بيولوجي: جرثومة أو فيروس مصمم خصيصاً يقتل جميع البشر. وإذا كانت قابلية نقل الفيروس عالية بما فيه الكفاية وفترة حضانته طويلة بما فيه الكفاية؛ فسيتمكن من إصابة الجميع قبل أن يدركوا وجوده ويتخذوا تدابير مضادة. وهناك حجة عسكرية لبناء سلاح بيولوجي كهذا حتى ولو لم يكن بوسعه قتل الجميع: إن أكثر أدوات يوم الهلاك فعالية هي تلك التي تجمع بين الأسلحة النووية والبيولوجية وغيرها من الأسلحة لزيادة فرص ردع العدو.

أسلحة الذكاء الاصطناعي

قد يشتمل مسار تكنولوجي ثالث يؤدي إلى انتحار البشرية على أسلحة ذكاء اصطناعي غبية نسبيّاً. لنفترض أن قوة عظمى تبني البلايين من الدرونات التي بحجم النحلة من الفصل الثالث وتستخدمها لقتل أي شخص باستثناء مواطنيها وحلفاؤها الذين يتم تحديدهم عن بعد ببطاقات هوية ذات ترددات الراديو كما هي الحال اليوم مع معظم منتجات السوبر ماركت. ويمكن توزيع هذه العلامات على جميع المواطنين على أساور أو غرسات عبر الجلد، كما في الجزء الذي تناول الأنظمة التوتاليتارية، هذا من شأنه أن يحفز قوة عظمى معارضة لبناء شيء مماثل. وعندما تندلع الحرب بطريق الخطأ، فإن جميع البشر سوف يُقتلون، حتى القبائل النائية غير المنخرطة في الحرب، لأنه لن يكون هناك أي شخص يرتدي كلا النوعين من الهوية. وبالجمع بين هذا وأجهزة يوم الهلاك النووية والبيولوجية؛ فإن هذا ستزداد فرصة انتحار البشرية.

ما الذي تريده أنت؟

أنت بدأت هذا الفصل مُفكِّراً في النقطة التي تريد أن يصل إليها سباق الذكاء AGI الحالي. الآن، وقد اكتشفنا مجموعة واسعة من السيناريوهات معاً، ما هي السيناريوهات التي تروق لك وتلك التي تعتقد أننا يجب أن نحاول تجنبها؟ هل لديك سيناريو مفضل بوضوح؟ اسمحوا لي ولزملائي القراء بمعرفة ذلك بالمشاركة على http://AgeOfAi.org، والانضمام إلى المناقشة!

من الواضح أنه لا ينبغي النظر إلى السيناريوهات التي ذكرناها على أنها قائمة شاملة لكل ما هو ممكن، والكثير منها متشعب التفاصيل بالفعل. ولكنني حاولت جاهداً أن أكون شاملاً، فأتطرق إلى المدى الكامل من التكنولوجيا المتقدمة إلى التكنولوجيا المنخفضة إلى اللا تكنولوجيا واصفاً جميع الآمال والمخاوف الرئيسية المعبر عنها في الأدب.

أحد أكثر الجوانب الممتعة في كتابة هذا الكتاب هو سماع رأي أصدقائي وزملائي في هذه السيناريوهات. وقد شعرت بالراحة عندما علمت أنه لا يوجد إجماع على الإطلاق، الشيء الوحيد الذي يتفق عليه الجميع هو أن الخيارات أكثر دقة مما قد يبدو في البداية. والأشخاص الذين يميلون إلى سيناريو واحد من السيناريوهات يعثرون على جانب (أو جوانب) مزعجة فيه. وبالنسبة إلي، هذا يعني أننا نحن البشر بحاجة إلى مواصلة هذه المحادثة حول أهدافنا المستقبلية وتعميقها، حتى نعرف إلى أي اتجاه نتوجه. والإمكانات المستقبلية للحياة في كوننا مذهلة جدّاً، لذلك دعونا لا نهدرها بالانجراف كسفينة بلا دقّة، جاهلين المكان الذي نقصده!

ما مدى ضخامة هذه الإمكانات المستقبلية؟ بغض النظر عن مدى تقدمنا التكنولوجي، فإن قدرة الحياة 3.0 على التحسن والانتشار عبر كوننا ستكون محدودة بقوانين الفيزياء - ما هي هذه الحدود القصوى خلال البلايين من السنين القادمة؟ هل يعج كوننا الآن بالحياة خارج كوكب الأرض، أم أننا وحدنا؟ ماذا يحدث إذا التقت مختلف الحضارات الكونية المتوسعة؟ سنتناول هذه الأسئلة الرائعة في الفصل التالي.

الخلاصة:

- قد ينتهي السباق الحالي نحو الذكاء AGI بمجموعة واسعة ورائعة من سيناريوهات ما سيحدث بعد آلاف السنين.
- قد يتعايش الذكاء الفائق بسلام مع البشر إما لأنه مجبر على (سيناريو السيّد المُستعبَد) أو لأنه «ذكاء اصطناعي ودود» يريد ذلك (يوتوبيا ليبرتارية، أو سيّد حامٍ، سيناريوهات الديكتاتور الخيّر، وحارس حديقة الحيوان).
- يمكن منع بناء الذكاء الفائق من خلال الذكاء الاصطناعي (سيناريو حارس البوابة)، أو البشر (سيناريو 1984)، أو بالنسيان المتعمد للتكنولوجيا (سيناريو الارتداد)، أو عن طريق عدم وجود حوافز لبنائه (سيناريو يوتوبيا المساواة).
- قد تنقرض البشرية ويحل محلها ذكاءات اصطناعية (سيناريوهات الفاتح والأحفاد) أو لا شيء (سيناريو التدمير الذاتي).
- لا يوجد إجماع على الإطلاق حول أي من هذه السيناريوهات مرغوب فيه، وكلها تنطوي على عناصر مرفوضة، وهذا يجعل من الأهمية بمكان مواصلة المحادثة حول أهدافنا المستقبلية وتعميقها، حتى لا ننجرف أو نتوجه عن غير قصد في اتجاه مؤسف.

الفصل 6

وقفنا الكوزموسي:
البليون سنة القادمة وما بعدها

تنتهي تكهناتنا بالوصول إلى حضارة فائقة، وهي توليفة من جميع الحياة في النظام الشمسي، وتعمل باستمرار على تحسين نطاقها وتوسيعه، وتمتد إلى الخارج من الشمس، محوِّلة ما ليس بحياة إلى عقل.

هانز مورافيك، **أطفال العقل**

بالنسبة إليّ، فإن الاكتشاف الأكثر إلهاماً على الإطلاق هو أننا بخسنا بشكل كبير من قدر إمكانات الحياة المستقبلية. فلا ينبغي أن تقتصر أحلامنا وتطلعاتنا على فترات تمتد على مدى قرن من الزمن يشوبها المرض والفقر والارتباك. وبدلاً من ذلك، وبفضل التكنولوجيا، فإن الحياة لديها القدرة على الازدهار لبلايين السنين، ليس فقط في نظامنا الشمسي، ولكن أيضاً في عالم أكبر بكثير وأكثر إلهاماً مما تصور أسلافنا. ولا حتى السماء هي الحد.

هذه أخبار مثيرة لنوعٍ Species يستوحي الإلهام من دفع الحدود عبر العصور. فالألعاب الأولمبية تحتفل بدفع حدود القوة والسرعة والرشاقة والتحمل. والعلم يحتفل بدفع حدود المعرفة والفهم. والأدب يحتفل بدفع حدود الإبداع الجميل أو الخبرات التي تثري الحياة. والعديد من الأفراد، من منظمات أو أمم- يحتفلون بزيادة الموارد والمقاطعات وطول العمر. وإذا أخذنا بعين الاعتبار هَوسَنا كبشر بالحدود، فمن الملائم أن الكتاب المحمي بعلامة تجارية والأفضل مبيعاً في جميع العصور هو كتاب **غينيس للأرقام القياسية العالمية** ***The Guinness Book of World Records***.

لذا، إذا كانت التكنولوجيا قادرة على تحطيم حدود الحياة المدركة منذ القدم، فما هي الحدود **القصوى**؟ كم من كوننا قد يغدو حيّاً؟ إلى أي مدى قد تصل الحياة وإلى متى

قد تستمر؟ كم من المواد يمكن للحياة الاستفادة منها، وكم من الطاقة والمعلومات والحوسبة يمكن استخلاصها؟ لا يتم تحديد هذه الحدود النهائية وفقاً لفهمنا، ولكن وفقاً لقوانين الفيزياء. ومن المفارقات أن هذا يجعل تحليل مستقبل الحياة على المدى الطويل أسهل منه من المستقبل القريب.

إذا ضغطنا تاريخنا الكوني الذي يبلغ 13.8 بليون عام في غضون أسبوع، فإن الدراما التي استمرت عشرة آلاف عام عبر الفصلين الأخيرين ستنتهي في أقل من نصف ثانية. وهذا يعني أنه على الرغم من أننا لا نستطيع التنبؤ بما إذا كان الذكاء سينفجر ولا بالكيفية التي سيتجلى فيها انفجار ذكاء وما سيكون عليه بعد ذلك، فإن كل هذا الاضطراب هو مجرد ومضة قصيرة في التاريخ الكوني الذي لا تؤثر تفاصيله في حدود الحياة القصوى. وإذا كانت حياة ما بعد الانفجار مهووسة مثلما هم البشر اليوم بدفع الحدود، فسوف تطور تكنولوجيا قادرة على الوصول إلى هذه الحدود فعليّاً - لأنها تستطيع ذلك. وفي هذا الفصل سنستكشف ماهية هذه الحدود، ومن ثم نلقي نظرة على ما قد يكون عليه مستقبل الحياة على المدى الطويل. ولما كانت هذه الحدود قائمة على فهمنا الحالي للفيزياء، يجب أن يُنظر إليها على أنها الحدود الأدنى للاحتمالات: فالاكتشافات العلمية المستقبلية قد توفر فرصاً للقيام بعمل أفضل.

لكن، هل نعرف حقّاً أن الحياة المستقبلية ستكون طموحة إلى هذه الدرجة؟ لا، نحن لا نعرف: ربما سنتكون راضياً مثل مدمن الهيروين أو شخص كسول يظل مستلقياً على الأريكة يشاهد إعادة مسلسل **مواكبة عائلة كارداشيان** ***Keeping Up with the Kardashians***. ولكن، هناك سبباً للشك في أن الطموح هو سمة عامة للحياة المتقدمة. وبغض النظر تقريباً عما تحاول تحقيق أقصى حدوده، سواء أكان ذكاء أم طول عمر أم معرفة أم تجارب مثيرة للاهتمام، فسوف يحتاج ذلك إلى موارد. لذلك، لديها حافز لدفع تكنولوجياتها إلى أقصى الحدود، لتحقيق أقصى استفادة من الموارد التي لديها. وبعد ذلك، فإن الطريقة الوحيدة لمزيد من التحسين هي الحصول على مزيد من الموارد، من خلال الانتشار في مناطق أكبر من الكوزموس.

أيضا، قد تنشأ الحياة بشكل مستقل في أماكن متعددة في الكوزموس الخاص بنا. وفي هذه الحالة تصبح الحضارات غير الطموحة ببساطة غير مهمة من الناحية السياسية، مع استيلاء أشكال الحياة الأكثر طموحاً على أجزاء أكبر من الوقف الكوزموسي في نهاية المطاف. ومن ثم، يُطبَّق الانتخاب (الانتقاء) الطبيعي Natural selection على نطاق كوزموسي، وبعد فترة، ستكون كل الحياة الموجودة تقريباً حياة طموحة. وباختصار، إذا كُنَّا مهتمين بالمدى الذي قد يصل إليه الكوزموس في نهاية المطاف، فيجب علينا دراسة حدود الطموح التي تفرضها قوانين الفيزياء. لنقم بهذا! فلنستكشف حدود ما يمكن القيام به بالموارد (مادة، طاقة، إلخ...) التي تتوفر لدينا في المجموعة الشمسية، ثم نعود إلى الكيفية التي نحصل بها على المزيد من الموارد عبر اكتشاف الكون واستيطانه.

الاستفادة القصوى من مواردك

بينما تبيع محلات السوبر ماركت وبورصات السلع الأساسية اليوم عشرات الآلاف من العناصر التي قد نسميها "موارد" Resources، فإن الحياة المستقبلية التي تصل إلى الحد التكنولوجي ستحتاج بشكل أساسي إلى مورد أساسي واحد: ما يسمى **المادة الباريونية** ***Baryonic matter***، وهذا يعني أي شيء يتكون من الذرات أو مكوناتها (الكواركات والإلكترونات). وأيّاً كان شكل هذه المادة، يمكن للتكنولوجيا المتقدمة إعادة ترتيبها في أي مواد أو أشياء مطلوبة، بما في ذلك محطات توليد الطاقة والحواسيب وأشكال الحياة المتقدمة. إذاً، فلنبدأ بدراسة القيود المفروضة على الطاقة التي تغذي الحياة المتقدمة، ومعالجة المعلومات التي تمكنها من التفكير.

بناء كرات دايسون

عندما يتعلق الأمر بمستقبل الحياة، فإن فريمان دايسون Freeman Dyson هو من مستشرفي المستقبل Visionary الذين يبعثون على الأمل. فقد تشرفتُ وسررتُ بمعرفته على مدى العقدين الماضيين. ولكن عندما التقيت به أول مرة، شعرت بالتوتر، كنت مجرد طالب صغير في مرحلة ما بعد الدكتوراه مع أصدقائي في غرفة الغداء بمعهد الدراسات المتقدمة Institute for Advanced Study في برينستون، وفجأة، دخل هذا الفيزيائي العالمي الذي كان يقضي الوقت مع آينشتاين وغودل وقدَّم نفسه، وسأل إذا كان يمكنه الانضمام إلينا! سرعان ما شعرنا بارتياح عندما شرح أنه يفضل تناول الغداء مع أشخاص يافعين على تناوله مع الأساتذة الكبار في السن المتجهمين. وعلى الرغم من أنه في الثالثة والتسعين من العمر وأنا أطبع هذه الكلمات، فإن فريمان لا يزال أصغر سنّاً من حيث الروح من معظم الذين أعرفهم. ويكشف البريق المشاكس في نظرته إلى أنه لا يهتم بالشكليات أو التسلسل الهرمي الأكاديمي أو الحكمة التقليدية. وكلما كانت الفكرة أكثر جرأة، كان متحمساً أكثر.

وعندما تحدثنا عن استخدام الطاقة، سخر من قلة طموحنا نحن البشر، مشيراً إلى أنه يمكننا تلبية جميع احتياجاتنا العالمية من الطاقة من خلال حصاد أشعة الشمس التي تضرب منطقة أصغر من 0.5% من الصحراء الكبرى. ولكن لماذا نتوقف عند هذا الحد؟ لماذا لا نلتقط جميع أشعة الشمس التي تصل إلى الأرض، ونترك معظمها يشع في الفضاء الخالي؟ لماذا لا نستخدم ببساطة جميع طاقة الشمس المُنتجة؟

ومستوحياً من رواية الخيال العلمي الكلاسيكية **صانع النجوم** ***Star Maker*** للمؤلف أولاف ستابليدون Olaf Stapledon، المنشورة في عام 1937، بحلقات من عوالم أثرية تدور حول نجمها الأم، فقد نشر فريمان دايسون وصفاً في عام 1960 لما صار يُعرف بكرة (مجال) دايسون Dyson sphere[1]، كانت فكرة فريمان تقوم على إعادة ترتيب كوكب المشتري إلى بيوسفير (محيط حيوي) Biosphere ضمن قشرة كروية، تحيط بالشمس، حيث يمكن لأحفادنا أن يزدهروا متمتعين بكتلة حيوية أكبر بمقدار 100 بليون ضعف، وبتريليون ضعف

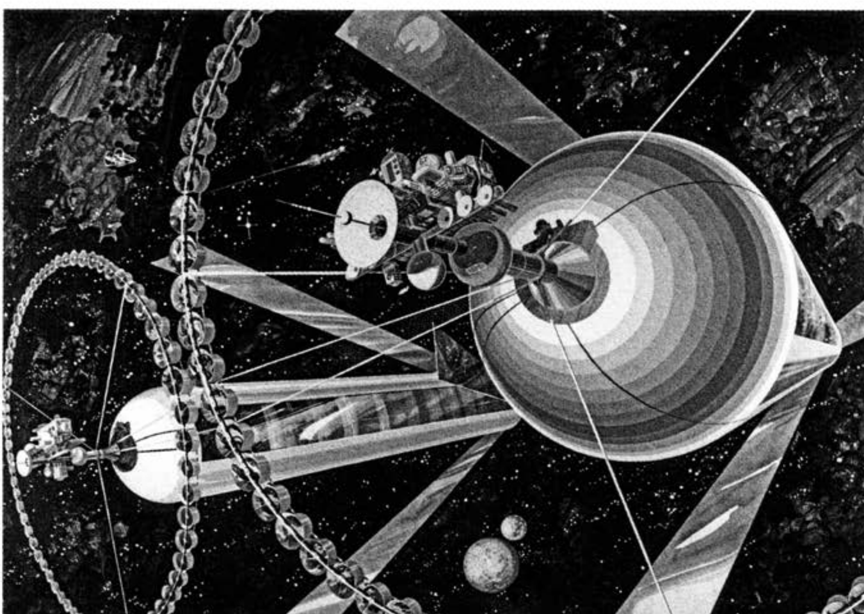
الطاقة التي تستهلكها البشرية اليوم.[2] وجادل في أن هذه هي الخطوة الطبيعية التالية: «يجب أن نتوقع أنه، خلال بضعة آلاف من السنين من دخوله مرحلة التنمية الصناعية، فإن أي نوعٍ Specied ذكي يجب أن يعيش في محيط حيوي يحيط بالكامل بنجمه الأم». وإذا كنت تعيش في كرة دايسون، لن يكون هناك ليل، وسترى ضوء الشمس دائماً منعكساً عن بقية البيوسفير، تماماً كما يمكنك في هذه الأيام رؤية أشعة الشمس منعكسة عن القمر خلال النهار. وإذا كنت ترغب في رؤية النجوم؛ فستذهب ببساطة إلى «الطابق العلوي» وتسترق النظر إلى الكون خارج كرة دايسون.

وإحدى طرق التكنولوجيا المنخفضة لإنشاء كرة دايسون جزئية هي وضع حلقة من الموائل في مدار دائري حول الشمس. ولتطويق الشمس تماماً، يمكنك إضافة حلقات تدور حول محاور مختلفة على مسافات مختلفة قليلاً لتجنب الاصطدامات. ولتجنب الإزعاج المتمثل بأن هذه الحلقات سريعة الحركة لا يمكن ربطها ببعضها البعض؛ مما يعقد النقل والاتصالات، يمكن للمرء بدلاً من ذلك أن يبني كرة دايسون واحدة -ثابتة ومتجانسة، حيث يتعادل سحب الجاذبية الداخلي للشمس بالضغط الخارجي الناتج من إشعاع الشمس - الفكرة الرائدة التي طرحها روبرت إل. فورورد Robert L. Forward وكولين ماكينس Colin McInnes. ويمكن بناء الكرة عن طريق إضافة المزيد من "الأقمار الثابتة" Statites تدريجيّاً: أقمار اصطناعية ثابتة تعادل جاذبية الشمس بضغط إشعاعي بدلاً من قوى الطرد المركزي. وتتناقص هاتان القوتان طرديّاً مع مربع المسافة إلى الشمس؛ مما يعني أنه إذا كان يمكن موازنتهما على مسافة واحدة من الشمس، فسيتم موازنتهما بسهولة من أي مسافة أخرى؛ مما يسمح بحرية الوقوف في أي مكان في نظامنا الشمسي. يجب أن تتألف الأقمار الثابتة من صفائح خفيفة الوزن جدّاً، تزن فقط 0.77 غرام لكل متر مربع، أي أقل بمئة ضعف من الورق، ولكن هذا من غير المرجح أن يعيق تطويرها. مثلاً، تزن طبقة من الغرافين Graphene (طبقة واحدة من ذرات الكربون مشكلة على نمط سداسي يشبه سلك الدجاج)، أي أقل بألف ضعف عن هذا الحد. وإذا صمّمنا كرة دايسون لتؤثر في ضوء الشمس بدلاً من امتصاصه؛ فستزداد الكثافة الكلية للارتداد المنعكس داخلها بشكل كبير؛ مما يزيد من ضغط الإشعاع وكمية الكتلة Mass التي يمكن دعمها في الكرة. كما أن العديد من النجوم الأخرى لها لمعان Luminosity أكبر بمقدار ألف ضعف، بل مليون ضعف شمسنا، ومن ثم فهي قادرة على دعم كرات دايسون ثابتة أثقل من المحيطة بشمسنا.

وإذا كان المطلوب هنا هو الحصول على كرة دايسون صلبة أثقل بكثير هنا في نظامنا الشمسي، فإن مقاومة جاذبية الشمس ستحتاج إلى مواد قوية جدّاً يمكنها تحمل ضغط أعلى بعشرات الآلاف من المرات من تلك الموجودة في قاعدة أطول ناطحات سحاب في العالم، دون تسييلها أو ربطها. ولكي تكون كرة دايسون طويلة العمر، ستحتاج إلى أن تكون ديناميكية وذكية، وأن تحدد موقعها وشكلها باستمرار استجابةً للاضطرابات وتفتح أحياناً فتحات كبيرة للسماح بمرور الكويكبات والمذنبات المزعجة دون حوادث. وبدلاً من ذلك، يمكن استخدام نظام الكشف والتطور للتعامل مع مثل هؤلاء المتسللين إلى النظام، وتفكيكها واستخدام مادتها على نحو أفضل.

وفي حالة بشر اليوم، فإن الحياة على أو داخل كرة دايسون ستكون في أحسن الأحوال مربكة، ومستحيلة في أسوأ الأحوال. ولكن هذا لن يمنع ازدهار أشكال الحياة البيولوجية أو غير البيولوجية هناك في المستقبل. ولن يلقي الشكل المداري أساساً أي جاذبية على الإطلاق، وإذا كنت تتجول في النمط الثابت من الكرة؛ فسيمكنك المشي فقط من الخارج (بعيداً عن الشمس) دون السقوط، بجاذبية أقل بعشرة آلاف مرة مما أنت معتاد عليها. فليس لديك مجال مغناطيسي (ما لم تبنِ مجالاً مغناطيسيّاً) يحميك من الجسيمات الخطرة من الشمس. والجانب الإيجابي لذلك، هو أن كرة دايسون بحجم مدار الأرض الحالي ستمنحنا مساحة أكبر بما يعادل 500 مليون ضعف لنعيش عليها.

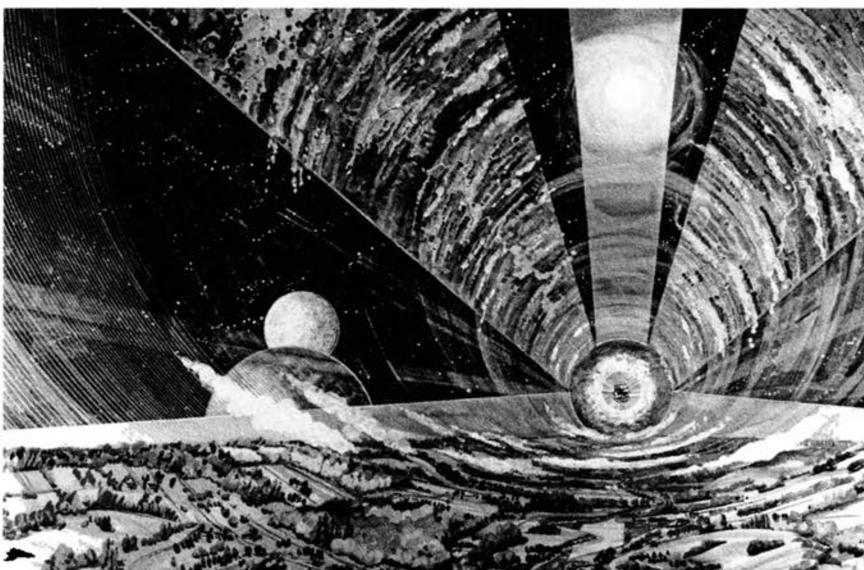
وإذا كانت هناك رغبة بالمزيد من الموائل البشرية الشبيهة بالأرض، فإن الخبر السار هو أن بناءها أسهل بكثير من بناء كرة دايسون. مثلاً، يعرض الشكلان 1.6 و 2.6 تصميماً للموائل الأسطوانية التي ابتكرها الفيزيائي الأمريكي جيرارد كيه. أونيل Gerard K. O'Neill، والتي تدعم نشوء جاذبية أرضية، وهناك درع تقي من الأشعة الكونية، ودورة ليل/نهار كل أربع وعشرين ساعة، وغلاف جوي Atmosphere ونظم إيكولوجية مثل الأرض. ويمكن لمثل هذه الموائل أن تدور بحرية داخل كرة دايسون، أو يمكن إضافة ملحقات بمتغيرات معدلة تلتحم بالكرة من الخارج.



الشكل 1.6: يمكن لزوج من أسطوانات أونيل التي يدور كل منهما بعكس الآخر أن توفر موائل بشرية مريحة شبيهة بالأرض إذا كانت تدور حول الشمس بطريقة تقابلها دائماً. وتوفر قوة الطرد المركزي الناتجة منهما جاذبية اصطناعية، وتعكس المرايا الثلاث القابلة للطي أشعة الشمس نحو الداخل على جدول يماثل ساعات اليوم والنهار الأربع والعشرين. أما الموائل الأصغر المرتبة في الحلقة، فهي متخصصة بالزراعة. الصورة إهداء من Rick Guidice/NASA.

بناء محطات طاقة أفضل

على الرغم من أن كرات دايسون تتميز بالكفاءة في استخدام الطاقة وفقاً للمعايير الهندسية الحالية، فإنها لا تقترب من دفع الحدود التي تحددها قوانين الفيزياء. وقد علمنا آينشتاين أننا إذا استطعنا تحويل الكتلة إلى طاقة بكفاءة 100%،* عندئذٍ فإن كمية من الكتلة m ستعطينا كمية من الطاقة E نحسبها وفقا لمعادلته الشهيرة $E = mc^2$، حيث c هي سرعة الضوء. وهذا يعني أنه لما كانت قيمة c ضخمة، فإن كمية صغيرة من الكتلة ستنتج كمية هائلة من الطاقة. وإذا كانت لدينا إمدادات وفيرة من ضديد المادة Antimatter (وهو ما لا ليس لدينا)، فعندئذ سيكون من السهل بناء محطة طاقة بكفاءة 100%: ببساطة صب ملعقة صغيرة من ضديد الماء في المياه العادية، وستطلق طاقة تعادل 200 ألف طن من مادة تي إن تي TNT، أي ما يعادل حصيلة قنبلة هيدروجينية نموذجية، تكفي لتزويد كامل احتياجات الطاقة في العالم لنحو سبع دقائق.

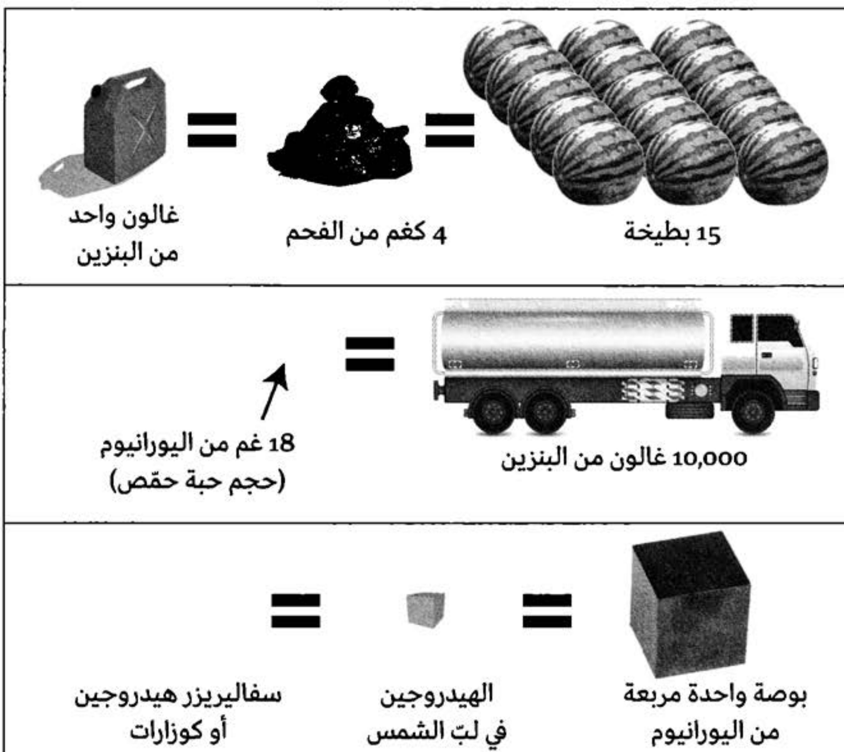


الشكل 2.6: منظر داخلي لإحدى أسطوانات أونيل O'Neill من الشكل السابق، إذا كان قطرها 6.4 كيلومتر وتدور مرة واحدة كل دقيقتين، فسيواجه الأشخاص الموجودون على السطح الجاذبية نفسها الظاهرة على الأرض. الشمس وراءك، لكنها تظهر أعلاه بسبب وجود مرآة خارج الأسطوانة التي تطوى في الليل. تمنع النوافذ محكمة الإغلاق الغلاف الجوي من الهروب من الأسطوانة. الصورة إهداء من Rick Guidice/NASA.

* إذا كنت تعمل في قطاع الطاقة، فقد تكون بدلاً من ذلك معتاداً على تحديد الكفاءة التي تمثل نسبة الطاقة المنبعثة في صورة مفيدة.

العملية	الفاعلية
هضم قطعة حلوى	%0.00000001
حرق الفحم	%0.00000003
حرق البنزين	%0.00000005
انشطار اليورانيوم - 235	%0.08
استخدام كرة (مجال) دايسون حتى تنطفئ الشمس	%0.08
اندماج الهيدروجين في الهيليوم	%0.7
محرك طاقة يتألف من ثقب أسود دوّار	%29
كرة (مجال) دايسون حول كوازار	%42
سفاليرايز Sphalierizer	%50
تبخير ثقب أسود	%90

الجدول 1.6: كفاءة تحويل الكتلة إلى طاقة قابلة للاستخدام نسبة إلى الحد النظري $E=mc^2$. كما هو موضح في النص، فإن الوصول إلى كفاءة تعادل 90% بإطعام الثقوب السوداء بالمادة وانتظار تبخرها هو مع الأسف أمر بطيء جدّاً فلا يكون مفيداً، كما أن تسريع العملية يقلل بشكل كبير من الكفاءة.



الشكل 3.6: يمكن للتقنية المتقدمة أن تستخلص طاقة من المادة أكثر بكثير مما نحصل عليه من الأكل أو الحرق، وحتى الاندماج النووي يستخرج طاقة أقل بمقدار 140 مرة من الحدود التي تحددها قوانين الفيزياء. وقد تكون محطات توليد الطاقة التي تستخدم السفاليرون Sphalerons أو الكوازارات أو الثقوب السوداء المتبخرة أفضل بكثير.

وعلى النقيض من ذلك، فإن أكثر طرقنا شيوعاً لتوليد الطاقة اليوم غير مفيدة بشكل يرثى له، كما هو موضح في الشكل 1.6 والشكل 3.6. وإن عملية هضم قطعة الحلوى هي كفؤة بما يعادل 0.00000001%. بمعنى أنها تُطلق مجرد جزء من عشرة تريليونات من الطاقة mc^2 التي تحوي عليها قطعة الحلوى. وإذا كانت معدتك فعالة حتى بنسبة 0.001%. فأنت لن تحتاج إلا إلى تناول وجبة واحدة فقط لبقية حياتك كلها. مقارنة بتناول الطعام، فإن حرق الفحم والبنزين أكثر فعالية بمقدار ثلاثة وخمسة أضعاف، على التوالي. اليوم، تعمل المفاعلات النووية بشكل أفضل جدّاً بفلق ذرات اليورانيوم بالانشطار النووي. ولكنها لا تزال تفشل في استخراج أكثر من 0.08% من طاقتها. إن المفاعل النووي الموجود في قلب الشمس هو أعلى كفاءة بمرتبة كاملة من المفاعلات التي بنيناها، إذ يستخرج 0.7% من الطاقة من الهيدروجين بدمجه مكوناً الهيليوم. لكن، حتى ولو وضعنا الشمس في كرة دايسون مثالية، فلن نتمكن أبداً من تحويل ما هو أكثر من نحو 0.08% من كتلة الشمس إلى طاقة يمكننا استخدامها، لأنه بمجرد أن تستهلك الشمس نحو عُشر وقودها الهيدروجيني؛ فستنتهي حياتها كنجم عادي، وتتمدد إلى عملاق أحمر، وتموت تدريجيّاً. وليست الحال أفضل بالنسبة إلى النجوم الأخرى أيضاً: يتراوح جزء الهيدروجين المُستهلك خلال فترة العمر الرئيسية من نحو 4% للنجوم الصغيرة جدّاً إلى نحو 12% بالنسبة للنجوم الأكبر. وإذا نجحنا في تحسين مفاعل الاندماج النووي الذي سيتيح لنا دمج 100% من جميع الهيدروجين المُتاح لنا، فإننا سنظل عالقين عند ذلك المستوى المنخفض بشكل محرج بنسبة 0.7% من عملية الاندماج. فأنّى لنا أن نحقق ما هو أفضل من ذلك؟

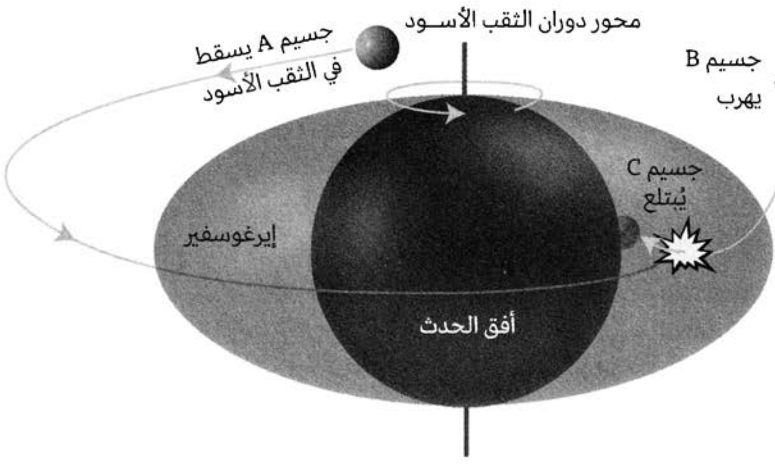
تبخير الثقوب السوداء

في كتابه تاريخ موجز للزمن ***A Brief History of Time*** اقترح ستيفن هوكينغ إنشاء محطة لتوليد الكهرباء من ثقب أسود.* قد يبدو أن هناك تناقضاً في هذا، فقد اعتقدنا لفترة طويلة أن الثقوب السوداء هي فخ لا يمكن لأي شيء، ولا حتى الضوء، الفرار منها. لكن، هوكينغ حسب الحسبة الشهيرة التي بيّنت أن تأثيرات الجاذبية الكمّيّة تجعلُ ثقباً أسود يعمل كجسم ساخن - كلما صغر كان أكثر سخونة - فينتج منه إشعاع حراري يعرف الآن بإشعاع هوكينغ Hawking radiation. وهذا يعني أن الثقب الأسود يفقد الطاقة تدريجيّاً ويتبخر، بعبارة أخرى، أيّاً كانت المادة التي ترميها في الثقب الأسود ستعود مرة أخرى في النهاية كإشعاع حراري، لذلك بحلول الوقت الذي يتبخر فيه الثقب الأسود تماماً، تكون مادتك قد تحولت إلى إشعاع بنسبة 100% تقريباً.**

* إذا لم يتم العثور على ثقب أسود مناسب من صنع الطبيعة في مكان قريب من الكون، فإنه يمكن إنشاء ثقب جديد عن طريق وضع الكثير من المادة في مساحة صغيرة بما فيه الكفاية.

** هذا تبسيط طفيف؛ لأن إشعاع هوكينغ يحوي أيضاً على بعض الجسيمات يصعب استخلاص ما هو مفيد منها. والثقوب السوداء الكبيرة ذات كفاءة تعادل 90% فقط؛ لأن نحو 10% من الطاقة تشع على شكل غرافيتونات Gravitons: جسيمات خجولة جدّاً يكاد يكون من المستحيل تقريباً اكتشافها، ناهيك عن استخلاص ما هو مفيد منها. ومع استمرار الثقب الأسود في الانكماش والتقلص، تنخفض الكفاءة أكثر فأكثر؛ لأن إشعاع هوكينغ يبدأ باحتواء نيوترينوات وجسيمات ضخمة أخرى.

وتكمن مشكلة استخدام تبخر الثقب الأسود كمصدر للطاقة في أنه ما لم يكن الثقب الأسود أصغر بكثير من حجم ذرة، فستكون العملية بطيئة جدّاً تستغرق وقتاً أطول من العصر الحالي لكوننا وتشع طاقة أقل من شمعة. وتتناقص القوة المنتجة مع مربع حجم الثقب الأسود، لذلك اقترح الفيزيائيان لويس كرين Louis Crane وشون ويستمورلاند Shawn Westmoreland استخدام ثقب أسود أصغر بنحو ألف مرة من البروتون، ويزن نحو وزن أكبر سفينة خاضت عباب البحر.[3] وكانت دوافعهما الرئيسية هي استخدام محرك الثقب الأسود لتشغيل مركبة فضائية (موضوع نعود إليه لاحقاً)، لذلك كانا أكثر اهتماماً بقابلية الحمل Portability من الكفاءة Efficiency. وقد اقترحا تغذية الثقب الأسود بضوء الليزر، دون التسبب في تحوُّل أي من الطاقة إلى المادة. وحتى لو تمكنت من إطعامه بالمادة بدلاً من الإشعاع، فإن ضمان كفاءة عالية يبدو أمراً صعباً: لجعل البروتونات تدخل مثل هذا الثقب الأسود الأصغر من حجمها بمقدار واحد على الألف، يجب أن تُقذف نحو الفتحة باستخدام آلة قوية مثل مصادم الهادرونات الكبير Large Hadron Collider (اختصاراً: المصادم LHC)، وزيادة طاقة البروتونات mc^2 مع ما لا يقل عن ألف ضعف طاقتها الحركية Kinetic (الحركة). ولما كان ما يعادل 10% على الأقل من تلك الطاقة الحركية فستنتقل إلى الغرافيتونات عندما يتبخر الثقب الأسود، ومن ثم فإننا نضع في الثقب الأسود طاقة أكبر مما يمكننا استخلاصه منها والاستفادة منه، وينتهي الأمر بكفاءة سلبية. ومما يزيد من تقييد احتمال إنشاء محطة طاقة ثقب السوداء، هو أننا ما زلنا نفتقر إلى نظرية راسخة للجاذبية الكمّيّة Quantum gravity التي قد تستند إليها حساباتنا - لكن عدم اليقين هذا بالطبع قد يعني أيضاً أن هناك آثاراً مفيدة للجاذبية الكمية لم تُكتشف بعد.



الشكل 4.6: يمكن استخلاص جزء من الطاقة الدورانية Rotational energy للثقب الأسود الدوار بإلقاء جسيم A بالقرب من الثقب الأسود وجعله ينقسم إلى جزء C يتم ابتلاعه وجزء B يهرب بطاقة أكبر من طاقة A الابتدائية.

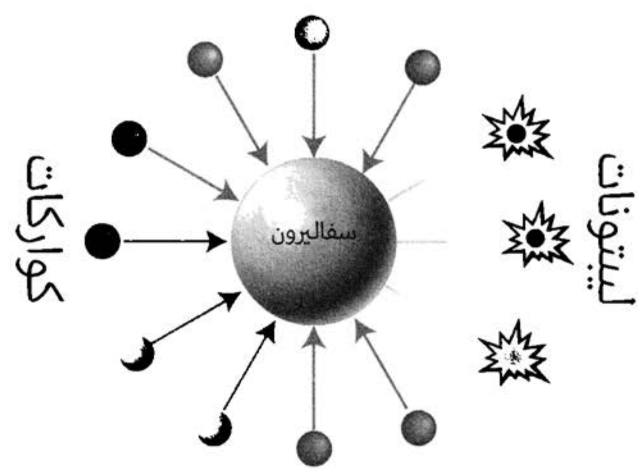
ثقوب سوداء دوّارة

لحسن الحظ، هناك طرق أخرى لاستخدام الثقوب السوداء كمحطات توليد للطاقة لا تتضمن الجاذبية الكمية أو غيرها من الفيزياء غير المفهومة. مثلاً، يدور العديد من الثقوب السوداء الموجودة بسرعة كبيرة، إذ تدور آفاق الحدث Event horizons الخاصة بها بسرعات تقارب سرعة الضوء، ويمكن استخلاص طاقة الدوران هذه. وأفق حدث الثقب الأسود هو المنطقة التي لا يمكن للضوء أن يهرب منها؛ لأن قوة جاذبيته كبيرة جدّاً. يوضح الشكل 4.6 كيف أنه إلى الخارج من أفق الحدث يكون للثقب أسود الدوار منطقة تُدعى **إيرغوسفير** (منطقة العمل) ***Ergosphere***، حيث يسحب الثقب الأسود الفضاء نحوه بسرعة كبيرة بحيث يتعذر معها على الجسيمات أن تقبع ساكنة دون أن تُسحب. وإذا رميت جسماً في الإيرغوسفير، فإنه سيكتسب سرعة بدورانه حول الثقب. ومع الأسف، سرعان ما سيبتلعه الثقب الأسود، ويختفي إلى الأبد عبر أفق الحدث، لذلك لن يفيدك ذلك إذا كنت تحاول استخلاص طاقة. ولكن، اكتشف روجر بنروز Roger Penrose أنه إذا أطلقت الجسم بزاوية ذكية وجعلته ينقسم إلى قطعتين كما يوضح الشكل 4.6. فيمكنك الإعداد لابتلاع قطعة واحدة في حين تفلت الأخرى من الثقب الأسود بكمية طاقة أكبر من طاقتها الابتدائية. وبعبارة أخرى، فقد نجحت في تحويل بعض الطاقة الدورانية للثقب الأسود إلى طاقة مفيدة يمكنك استخدامها. وبتكرار هذه العملية عدة مرات، يمكنك استخلاص **جميع** طاقة الثقب الأسود الدورانية حتى يتوقف عن الدوران ويختفي الإيرغوسفير. وإذا كان الثقب الأسود الأوليّ يدور بأسرع ما تسمح به الطبيعة، بأفق حدث يتحرك جوهريّاً بسرعة الضوء، ستسمح لك هذه الاستراتيجية بتحويل 29% من كتلته إلى طاقة. ولا يزال هناك قدر كبير من عدم اليقين حول مدى سرعة دوران الثقوب السوداء في سماء الليل، ولكن يبدو أن العديد من الثقوب السوداء المدروسة أفضل من غيرها التي تدور بسرعة كبيرة: بين 30% و 100% من الحد الأقصى المسموح به. والثقب الأسود الضخم في منتصف مجرتنا (الذي تعادل كتلته أربعة ملايين ضعف شمسنا) يبدو كأنه يدور، لذلك حتى ولو تمكنا من تحويل 10% فقط من كتلته إلى طاقة مفيدة، فإن ذلك سيوفر ما يعادل 400000 شمس حُوّلت إلى طاقة بكفاءة تعادل 100%، أو ما يقرب الطاقة التي سنحصل عليها من كرات (مجالات) دايسون نحو 500 المقامة حول مليون الشمس على مدى بلايين السنين.

كوازارات

هناك استراتيجية أخرى مثيرة للاهتمام لاستخلاص الطاقة ليس من الثقب الأسود نفسه، ولكن من المادة التي تقع فيه. وجدت الطبيعة طريقة للقيام بهذا كله بمفردها: الكوازار Quasar. فبينما يدور الغاز بالقرب من ثقب أسود، مكوناً قرصاً على شكل

قرص بيتزا تتناثر أجزاؤه الداخلية تدريجيّاً، يسخن الغاز جدّاً وتنبعث منه كميات كبيرة من الإشعاع. وبينما يسقط الغاز نحو الثقب، فإنه يتسارع؛ مما يحول طاقة الوضع Potential energy إلى جاذبية حركة، تماماً مثل القفز بالمظلات. وتصبح الحركة أكثر فوضوية بشكل تدريجي، إذ يعمل الاضطراب Turbulence المعقد على تحويل الحركة المنسقة لبقعة الغاز إلى حركة عشوائية على نطاقات أصغر يتزايد صغرها، حتى تبدأ الذرات بالاصطدام ببعضها البعض عند ارتفاع درجة الحرارة. هذه الحركة العشوائية هي بالضبط ما يعنيه أن كون الشيء حارّاً، وهذه التصادمات العنيفة تحول طاقة الحركة إلى إشعاع. ومن خلال بناء كرة (مجال) دايسون حول ثقب أسود بكامله، على مسافة آمنة، يمكن التقاط هذه الطاقة الإشعاعية واستخدامها. وكلما زادت سرعة دوران الثقب الأسود، زادت فعالية هذه العملية، إذ يوفر ثقبٌ أسود دوار طاقةً بكفاءة ضخمة بمعدل 42%.* وفي حالة ثقب أسود ذي كتلة تعادل كتلة نجم، فإن معظم الأشعة الصادرة ستكون أشعة سينية X-rays، في حين أن الثقوب السوداء الفائقة الموجودة في مراكز المجرات، ستكون أغلب الأشعة الصادرة عنها في نطاق الأشعة تحت الحمراء Infrared، والضوء المرئي والأشعة فوق البنفسجية Ultraviolet.



الشكل 5.6: وفقاً للنموذج القياسي (المعياري) لفيزياء الجسيمات، فقد تتحد تسعة كواركات من نكهة Flavor ودوران مغزلي Spin مناسبين وتتحول إلى ثلاثة لبتونات من خلال حالة وسيطة تسمى السفاليرون Sphaleron. والكتلة الكلية للكواركات (جنباً إلى جنب مع طاقة جسيمات الغلون Gluon المرافقة لها) أكبر بكثير من كتلة اللبتونات، لذلك ستطلق هذه العملية الطاقة، المُشار إليها في الشكل بالومضات.

* بالنسبة إلى عشاق دوغلاس آدمز Douglas Adams، لاحظ أن هذا سؤال أنيق يقدم الإجابة عن سؤال الحياة والكون وكل شي. وبتعبير أدق، تبلغ الكفاءة 42% ≈ $1 - 1/\sqrt{3}$.

وبمجرد أن تنفد من الوقود لإطعام ثقبك الأسود، يمكنك الانتقال إلى استخلاص طاقتها الدورانية كما ناقشنا آنفا.* بالطبع، فقد وجدت الطبيعة بالفعل طريقة للقيام بذلك جزئيّاً: تعزيز الإشعاع الناتج من تراكم الغاز من خلال عملية مغناطيسية المعروفة بآلية بلاندفورد-زنايك Blandford- Znajek mechanism. ربما أمكن استخدام التكنولوجيا لزيادة تحسين كفاءة استخلاص الطاقة بما يتجاوز 24%، وذلك بالاستخدام الذكي للمجالات المغناطيسية أو مكونات أخرى.

سفاليرونات

هناك طريقة أخرى معروفة لتحويل المادة إلى طاقة لا تتضمن ثقوباً سوداء على الإطلاق: **عملية السفاليرون *Sphaleron***. هذه العملية تدمر الكواركات وتحولها إلى لبتونات Leptons: الإلكترونات Electrons وأبناء عمومتها الأثقل، جسيمات الميون Muon والتاو Tau، النيوترينوات Neutrinos أو ضديداتها.[4] كما يتضح في الشكل 5.6، فإن النموذج القياسي (المعياري) لفيزياء الجسيمات Standard model of particle physics يتنبأ بتسعة كواركات -بنكهات Flavor ودوران مغزلي Spin مناسبين- يمكنها أن تتحد وتتحول إلى ثلاثة لبتونات من خلال حالة وسيطة تسمى حالة السفاليرون Sphaleron، ولما كانت المدخلات تزن أكثر من الناتج، يتحول فرق الكتلة إلى طاقة وفقاً لمعادلة آينشتاين $E = mc^2$.

ومن ثم، فقد تتمكن الحياة الذكية في المستقبل من بناء ما أسميه جهاز سفاليرايز Sphalerizer: مُوّلد طاقة يعمل كمحرك ديزل بقوة مضاعفة. ويضغط محرك الديزل التقليدي مزيجاً من الهواء وزيت الديزل إلى أن ترتفع درجة الحرارة بدرجة تكفي لإشعاله وحرقه تلقائيّاً، وبعد ذلك يتمدد الخليط الساخن وينجز أعمالاً مفيدة، مثلاً دفع المكبس Piston. وكتلة ثاني أكسيد الكربون وغازات الاحتراق الأخرى هو أقل بنحو 0.0000005% مما كان في المكبس في البداية، وهذا الاختلاف في الكتلة يتحول إلى طاقة حرارة تدفع المحرك. وسيضغط جهاز سفاليرايز المادة العادية إلى بضعة كوادريليون درجة، ثم يسمح لها بالتمدد مرة أخرى فتبرد بمجرد قيام السفاليرونات بعملها.** نحن نعرف

* إذا أطعمت الثقب الأسود من خلال وضع سحابة غاز تدور حوله ببطء في الاتجاه نفسه، فسيدور هذا الغاز بشكل أسرع من أي وقت مضى إذ يتم سحبه وابتلاعه؛ مما يعزز دوران الثقب الأسود، تماماً مثل تدور متزلجة على الجليد بشكل أسرع عندما تجذب ذراعيها نحوها. وقد يحافظ هذا على دوران الثقب إلى أقصى حد؛ مما يمكّنك من استخراج نحو 42% من طاقة الغاز، ومن ثم 29% من الباقي، للحصول على إجمالي كفاءة قدره 42% + (42%-1) × 29% ≈ 59%.

** يجب أن يكون حارّاً بدرجة كافية لإعادة توحيد القوى الكهرومغناطيسية Electromagnetic forces والضعيفة Weak forces، وهو ما يحدث عندما تتحرك الجسيمات بأعلى سرعة عندما يتم تسريعها بمقدار 200 بليون فولت في مصادم جسيمات.

بالفعل نتيجة هذه التجربة؛ لأن كوننا المبكر قام بها نيابة عنَّا قبل نحو 13.8 بليون سنة، عندما كان حارّاً: بتحويل نحو 100% من المادة إلى طاقة، مع بقاء أقل من جزء من البليون من الجسيمات التي تتشكل منها المادة العادية: الكواركات والإلكترونات. لذا، فهو يشبه محرك الديزل، غير أنه أكثر كفاءة ببليون ضعف! الميزة الأخرى هي أنك لست بحاجة إلى أن تكون دقيقاً في اختيار ما ستغذيه به، فهو يعمل بأي شيء مصنوع من الكواركات، أي بأي مادة طبيعية.

وبفعل هذه العمليات مرتفعة درجة الحرارة، أنتج كوننا اليافع إشعاعاً (الفوتونات Photons والنيوترينوات) يزيد على تريليون ضعف المادة (الكواركات والإلكترونات التي تجمعت في وقت لاحق في شكل ذرات). وخلال 13.8 بليون سنة منذ ذلك الحين، حدث فصل كبير، إذ تركزت الذرات في المجرات والنجوم والكواكب، في حين صُبَّت معظم الفوتونات في الفضاء بين المجرات مُشكِّلة إشعاع الخلفية الكونية الميكروي (إشعاع خلفية الكون) Cosmic microwave background radiation المُستخدم لرسم صور من كوننا اليافع. ومن ثم، فإن أي شكل من أشكال الحياة المتقدمة التي قد تعيش في مجرة ما أو أي تركيز آخر من المادة قد يحول معظم مادته المتاحة إلى طاقة، بإعادة نسبة المادة إلى القيمة الصغيرة نفسها التي نشأت من كوننا المبكر من خلال إعادة توليد تلك الظروف الكثيفة الساخنة لفترة موجزة داخل جهاز سفاليرايز. ولحساب مدى كفاءة جهاز سفاليرايز فعلي، يحتاج المرء إلى أن يأخذ باعتبار تفاصيل أساسية في العملية: مثلاً، ما مدى الحاجة إلى منع جزء كبير من الفوتونات والنيوترونات من التسرب أثناء مرحلة الضغط؟ ولكن ما يمكننا قوله بالتأكيد، هو أن آفاق الطاقة في مستقبل الحياة أفضل بكثير مما تسمح به تكنولوجيتنا الحالية. ولم نتمكن حتى من بناء مفاعل اندماج Fusion reactor، ولكن يجب أن تكون التكنولوجيا المستقبلية قادرة على العمل بأفضل مما لدينا بعشرة وربما بمئة ضعف.

بناء حواسيب أفضل

إذا كان تناول العشاء أسوأ بعشرة بلايين مرة من الحد المادي للطاقة؛ فما مدى كفاءة الحواسيب الحالية؟ أسوأ من ذلك العشاء، كما سنرى الآن.

غالباً ما أقدم صديقي وزميلي سيث لويد Seth Lloyd على أنه الشخص الوحيد في المعهد MIT الذي يمكن القول إنه مجنون مثلي. وبعد قيامه بعمل رائد على الحواسيب الكمّيّة Quantum computers، استمر بكتابة كتاب يجادل به في أن عالمنا بكامله هو حاسوب كمّيّ. وغالباً ما نلتقي بعد ساعات العمل، ولم أكتشف بعد موضوعاً لا يُدلي برأي مثير بخصوصه. مثلاً، كما ذكرت في الفصل الثاني. لديه الكثير ليقوله عن الحدود النهائية للحوسبة. ففي ورقة علمية شهيرة نُشرت في عام 2000، بيّن أن سرعة الحوسبة محدودة

بالطاقة: إن إجراء عملية منطقية أولية Elementary logical operation في زمن يساوي T يتطلب طاقة تعادل في المتوسط $E= h\backslash 4T$، حيث h هي الكمية الفيزيائية الأساسية المعروفة بثابت بلانك Planck's constant، هذا يعني أن حاسوباً يبلغ وزنه كيلوغرام واحد يمكنه القيام بأكثر من 5×10^{50} عملية في الثانية على الأكثر - وهذا حجم ضخم يعادل 36 ضعف الحاسوب الذي أطبع عليه هذه الكلمات. وسنصل إلى هذا المستوى خلال بضعة قرون إذا استمرت القوة الحوسبية بالتضاعف كل عامين، كما اكتشفنا في الفصل الثاني. كما بيّن لويد أن الحاسوب الذي يبلغ وزنه كيلوغرام واحد يمكنه تخزين ما لا يقل عن 10^{31} بت، وهو أفضل ببليون بليون ضعف أي حاسوب محمول.

ولويد هو أول من يعترف بأن بلوغ هذه الحدود في الواقع قد يكون تحدّياً حتى بالنسبة إلى حياة خارقة الذكاء، لأن ذاكرة "الحاسوب" الأقصى الذي يزن كيلوغرام واحد تشبه انفجاراً نووياً حرارياً أو جزءاً صغيراً من الانفجار الكبير Big Bang. ولكنه متفائل بأن القيود العملية ليست بعيدة عن الوصول إلى الحدود النهائية. في الواقع، فقد عملت النماذج الأولية للحاسوب الكمّيّة بالفعل على تقليص ذاكرتها من خلال تخزين بت واحد لكل ذرة، وتوسيع نطاق ذلك سيتيح تخزين نحو 10^{25} بت/كغم - أفضل بتريليون ضعف من الحاسوب المحمول. وإضافة إلى ذلك، فإن استخدام الإشعاع الكهرومغناطيسي للتواصل بين هذه الذرات من شأنه أن يتيح نحو 5×10^{40} عملية في الثانية الواحدة - 31 ضعف وحدة المعالجة المركزية CPU في حاسوبي.

وباختصار، فإن إمكانات الحياة المستقبلية لحوسبة الأشياء وإدارتها أمر محير حقّاً: من حيث المرتبة، فإن أفضل الحواسيب العملاقة اليوم هي أبعد ما تكون عن الحاسوب الأقصى ذي الكيلوغرام الواحد، مقارنة بإشارة الانعطاف الوامضة في سيارة، وهو جهاز يخزن مجرد جزء بت واحد فقط من المعلومات؛ مما يؤدي إلى تشغيله وإيقافه مرة واحدة في الثانية.

موارد أخرى

من منظور الفيزياء، كل ما قد ترغب الحياة المستقبلية في تكوينه - من الموائل والآلات إلى أشكال الحياة الجديدة - هو ببساطة جسيمات أولية مرتبة بطريقة معينة. فمثلما الحوت الأزرق هو إعادة ترتيب لعوالق الكريل Krill، والكريل إعادة ترتيب لعوالق أخرى، فإن نظامنا الشمسي بكامله هو ببساطة إعادة ترتيب الهيدروجين خلال 13.8 بليون عام من التطور الكوني: أعادت الجاذبية ترتيب الهيدروجين إلى النجوم التي بدورها أعادت ترتيب الهيدروجين في ذرات أثقل، وبعد ذلك إعادت الجاذبية ترتيب هذه الذرات إلى كوكبنا الذي أعاد ترتيب العمليات الكيميائية والبيولوجية لهذه الذرات في شكل حياة.

الحياة المستقبلية التي تكون قد وصلت إلى هذه الحدود التكنولوجية سيكون بإمكانها إعادة ترتيب هذه الجسيمات بسرعة أكبر وبكفاءة أكبر: أولاً، باستخدام

قوتها الحاسوبية لتحديد الطريقة الأكثر كفاءة، ثم استخدام الطاقة المتاحة لها لإدارة عملية إعادة ترتيب المادة. فقد رأينا كيف يمكن تحويل المادة إلى كل من الحواسيب والطاقة، لذلك فهي المصدر الأساسي الوحيد الذي نحن بحاجة إليه.* وبمجرد أن تصطدم الحياة في المستقبل بالقيود المادية على ما يمكن أن تحققه باستخدام المادة، فلا يتبقى سوى طريقة واحدة لعمل المزيد: الحصول على المزيد من المادة. والطريقة الوحيدة للقيام بذلك هي من خلال انتشارنا عبر كوننا، ضاربين عميقاً في الفضاء!

كسب الموارد من خلال الاستيطان الكوني

ما مدى عظم وقفنا الكوزموسي؟ على وجه التحديد، ما هي الحدود العليا التي تضعها قوانين الفيزياء على مقدار المادة التي يمكن للحياة الاستفادة منها في المحصلة؟ إن وقفنا الكوزموسي كبير لدرجة مُحيرة، بالطبع. ولكن كم هي كبيرة بالضبط؟ يسرد الجدول 2.6 بعض الأرقام الأساسية. فقد صار كوكبنا حاليّاً ميتاً بنسبة 99.999999%، بمعنى أن هذا الجزء من مادته ليس جزءاً من محيطنا الحيوي، ولا يقدم شيئاً مفيداً للحياة بخلاف توفير الجاذبية والحقل المغناطيسي، هذا يرفع من احتمال أنه في يوم من الأيام سيُستَخدم مئة مليون ضعف ذلك في دعم الحياة بفاعلية. وإذا استطعنا استخدام جميع المادة في نظامنا الشمسي (بما في ذلك الشمس) على النحو الأمثل؛ فسنكون أفضل حالا بمليون مرة. واستيطان مجرتنا سوف يؤدي إلى نمو مواردنا إلى تريليون ضعف.

المنطقة	الجسيمات
محيطنا الحيوي	10^{43}
كوكبنا	10^{51}
نظامنا الشمسي	10^{57}
مجرتنا	10^{69}
مدى سفرنا بنصف سرعة الضوء	10^{75}
مدى سفرنا بسرعة الضوء	10^{76}
كوننا	10^{78}

الجدول 2.6: العدد التقريبي لجسيمات المادة (البروتونات والنيوترونات) التي يمكن للحياة في المستقبل أن تطمح إلى الاستفادة منها.

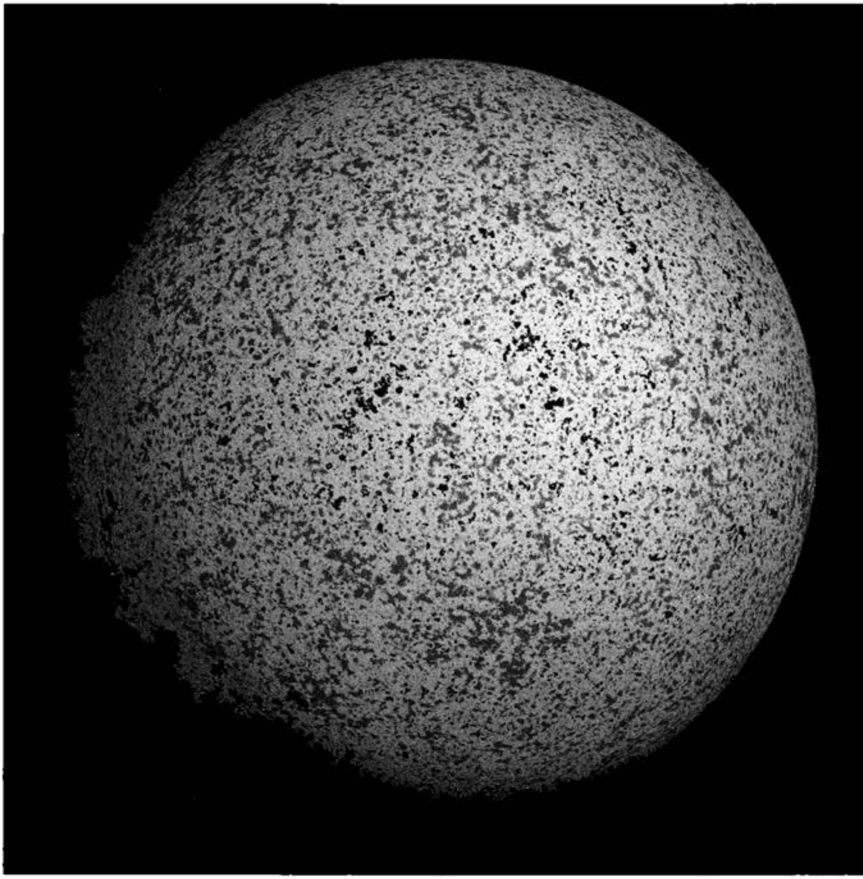
* ناقشنا آنفاً فقط المادة المصنوعة من الذرات. وهناك نحو ستة أضعافها من المادة المعتمة. لكنها مراوغة جدّاً ويصعب اقتناصها، في العادة تطير عبر الأرض مباشرة وتخرج من الجانب الآخر، لذلك يبقى لنا أن نرى ما إذا كان من الممكن للحياة المستقبلية اقتناصها واستخدامها.

إلى أي مدى يمكن أن تصل؟

قد تعتقد أنه يمكننا الحصول على موارد لانهائية باستيطان أكبر عدد ممكن من المجرات الأخرى إذا كنا صبورين بما فيه الكفاية، ولكن هذا ليس ما يقترحه علم الكونيات الحديث! نعم، قد يحوي الفضاء بحد ذاته عدداً لا نهائيّاً من المجرات والنجوم والكواكب - بالطبع، هذا ما تتنبأ به أبسط نمذجات **التوسع *Inflation***، النموذج الفكري العلمي السائد حول ما ولّدَه الانفجار الكبير منذ 13.8 بليون سنة. ولكن، وحتى لو كان هناك العديد من المجرات، يبدو أننا لا نستطيع أن نصل إلّا إلى عدد مُنتهٍ منها: يمكننا أن نرى نحو 200 بليون مجرة ونستوطن عشرة بلايين على الأكثر.

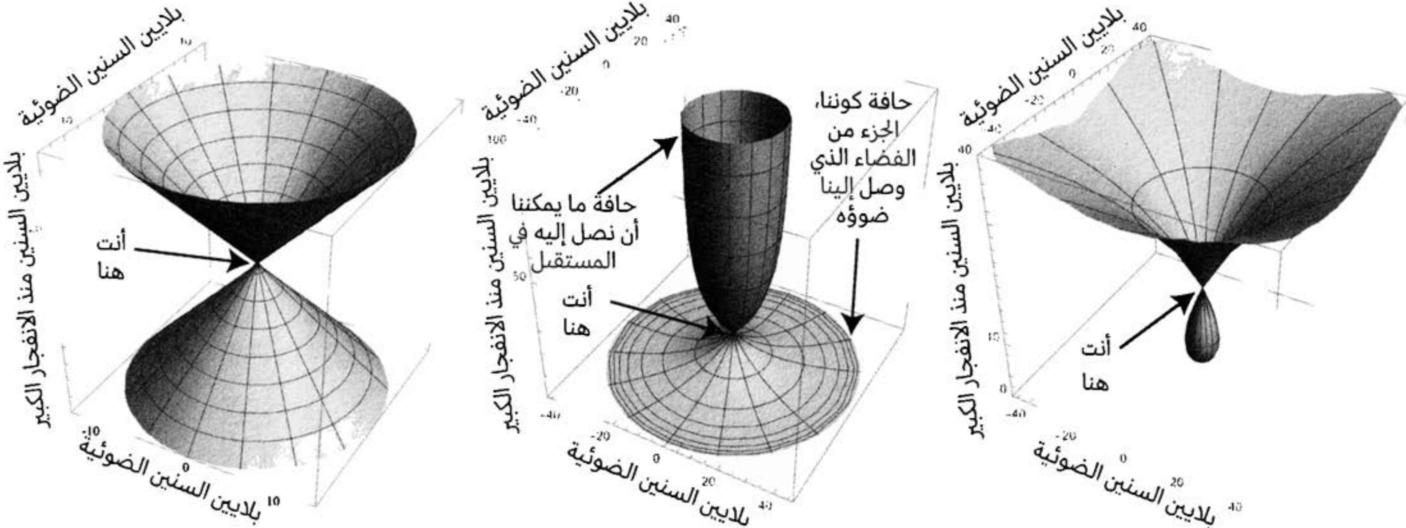
ما يحدّنا هو سرعة الضوء: سنة ضوئية واحدة (نحو عشرة تريليونات كيلومتر) في السنة. ويعرض الشكل 6.6 الجزء من الفضاء الذي وصلنا منه الضوء حتى الآن خلال 13.8 بليون عام منذ الانفجار الكبير، وهي منطقة كروية تُعرف بـ "كوننا الذي يمكن رصده" Our observable Universe أو ببساطة "كوننا". وحتى إذا كان الفضاء لا منتهياً، فإن كوننا يحتوي على نحو 10^{78} ذرة "فقط". وإضافة إلى ذلك، فإن نحو 98% من كوننا هو "للنظر وليس للمس"، بمعنى أنه يمكننا رؤيته لكننا لن نصل إليه أبداً حتى ولو سافرنا بسرعة الضوء، لماذا؟ بعد كل شيء، فإن الحد الأقصى للمدى الذي يمكننا رؤيته ينبع ببساطة من حقيقة أن كوننا ليس قديماً إلى ما لا نهاية، بحيث إنه لم يُتاح للضوء البعيد زمن كافٍ للوصول إلينا، لذا ألا ينبغي أن نكون قادرين على السفر بشكل اعتباطي إلى أيٍّ من المجرات البعيدة إذا لم تكن هناك حدود على مقدار الوقت الذي يمكننا قضاؤه في الطريق؟

التحدي الأول هو أن كوننا يتوسع؛ مما يعني أن جميع المجرات تقريباً تبتعد عنا، لذا فإن استيطان المجرات البعيدة يكاد تكون أشبه بلعبة الركض واللحاق. والتحدي الثاني هو أن هذا التوسع الكوزموسي يتسارع، بسبب الطاقة المعتمة الغامضة التي تشكل نحو 70% من عالمنا. ولفهم الكيفية التي يتسبب بها ذلك في حدوث المشكلة، تخيل أنك تدخل منصة قطار وترى أن القطار يسير ببطء بعيداً عنك، ولكن مع ترك باب من الأبواب مفتوحاً يدعوك للركوب، إذا كنت سريعاً ومجازفاً، فهل يمكنك اللحاق بالقطار؟ نظراً لأن القطار في النهاية سيكون أسرع من سرعة ركضك، فإن الإجابة تعتمد بوضوح على مدى بُعدك عن القطار في البداية: إذا تجاوز الأمر المسافة الحرجة، فلن تلحق به أبداً. نحن نواجه الموقف نفسه ونحاول صيد تلك المجرات البعيدة التي تتسارع مبتعدة عنا: حتى ولو تمكنا من السفر بسرعة الضوء، ستظل جميع المجرات التي هي أبعد عنا بنحو 17 بليون سنة ضوئية بعيدة عن مدانا إلى الأبد - ونسبتها تزيد على 98% من المجرات في كوننا.



الشكل 6.6: كوننا، أي المنطقة الكروية للفضاء التي أتيح للضوء فيها زمن كافي للوصول إلينا (في الوسط) خلال 13.8 بليون سنة منذ الانفجار الكبير. وتُظهر الأنماط صوراً لكوننا الوليد ملتقطة بواسطة قمر بلانك الاصطناعي، أي عندما كان عمره 400000 عام، وكان يتكون من بلازما حارة تقارب حرارة سطح الشمس. ربما يمتد الفضاء ما وراء هذه المنطقة، وفي كل عام يتبدّى لنا قدر جديد من المادة.

لكن انتظر: ألم تذكر نظرية النسبية الخاصة لآينشتاين أنه لا يوجد شيء يمكن أن يسافر بسرعة أكبر من سرعة الضوء؟ إذاً، كيف يمكن للمجرات أن تتفوق على شيء يسير بسرعة الضوء؟ الجواب هو أن نظرية النسبية العامة لآينشتاين تنسخ النسبية الخاصة، عندما يكون الحد الأقصى للسرعة أكثر حرية: لا شيء يمكن أن يسافر بسرعة أكبر من سرعة الضوء عبر الفضاء، ولكن الفضاء حرٌّ في أن يتوسع بالسرعة التي يريدها. فقد أعطانا آينشتاين أيضاً طريقة رائعة لتصور حدود السرعة هذه من خلال عرض الزمن باعتباره البعد الرابع في الزمكان Spacetime (انظر الشكل 7.6، حيث احتفظت بالأشياء في فضاء ثلاثي الأبعاد عن طريق حذف أحد أبعاد الفضاء الثلاثة). mإذا لم يكن الفضاء آخذا بالتوسع، فستكون خطوط أشعة الضوء مائلة بمقدار 45 درجة عبر الزمكان، وعندئذ ستكون المناطق التي



الشكل 7.6: في مخطط الزمكان، يكون الحدث عبارة عن نقطة تُرمَّز إحداثياتها الأفقية والعمودية لأين ومتى وقع الحدث، على التوالي. وإذا لم يتوسع المكان (اليسار)، فإن مخروطين اثنين يحددان أجزاء من الزمكان الذي نحن على الأرض (الطرف المدبب) نتأثر بها (المخروط السفلي) وقد يكون لنا تأثير فيها (المخروط العلوي)؛ لأن التأثيرات السببية لا يمكن أن تسافر أسرع من الضوء، الذي يسافر مسافة سنة ضوئية واحدة لكل سنة. وتصبح الأمور أكثر إثارة للاهتمام عندما يتمدد الفضاء (اليمين). ووفقاً للنموذج القياسي لعلم الكونيات، لا يمكننا رؤية سوى جزء منته من الزمكان والوصول إليه، وحتى لو كان الفضاء لامتنهياً. في الشكل الأوسط الذي يذكرنا بكأس زجاجية، نستخدم الإحداثيات التي تخفي توسع الفضاء، بحيث يمكن تمثيل حركات المجرات البعيدة بمرور الوقت بخطوط عمودية. ومن النقطة التي نرى بها الأشياء حاليّاً، أي بعد 13.8 بليون سنة من الانفجار الكبير، تكون أشعة الضوء التي أتيح لها الوقت للوصول إلينا فقط هي تلك المنطلقة من عند قاعدة الكأس الزجاجية، وحتى لو سافرنا بسرعة الضوء؛ فلا يمكننا الوصول إلى المناطق الواقعة إلى الخارج من الجزء العلوي جزء من الكأس الزجاجية الذي يحتوي على نحو عشرة بلايين مجرة. وفي الصورة إلى اليمين، الشبيهة بقطرة ندى تحت زهرة، نستخدم إحداثيات مألوفة حيث نشاهد الفضاء وهو يتوسع. يؤدي هذا إلى تشوه القاعدة الزجاجية إلى شكل قطرة؛ ندى نظراً لأن المناطق الموجودة عند حواف ما يمكننا رؤيته كانت قريبة جدّاً من بعضها البعض في وقت مبكر من الزمن.

يمكننا رؤيتها والوصول إليها الآن ومن هنا تقع ضمن شكل مخروطي. وبينما انتهى الانفجار الكبير مخروطنا السابق قبل 13.8 بليون سنة، فإن مخروط ضوئنا في المستقبل سوف يتوسع إلى الأبد؛ مما يتيح لنا الوصول إلى أبعاد غير محدودة من الوقف الكوزموسي. في المقابل، تُظهر الصورة الوسطى من هذا الشكل التوضيحي أن الكون المتسع ذا الطاقة المعتمة (الذي يبدو أنه الكون الذي نعيش فيه) يشوه مخاريطنا الضوئية إلى شكل يشبه كأساً زجاجية؛ مما يحد -إلى الأبد- من عدد المجرات التي يمكننا استيطانها إلى نحو عشرة ملايين مجرة.

وإذا كان هذا الحد يجعلك تشعر برهاب الاحتجاز الكوزموسي Cosmic claustrophobia، فدعني أخفف عنك باحتمال وجود ثغرة: حساباتي تفترض أن الطاقة المعتمة تظل ثابتة مع مرور الوقت، بما يتفق مع ما تقترحه القياسات الأخيرة. ولكن، ليس لدينا بعد أي فكرة عن ماهية الطاقة المعتمة حقّاً؛ مما يترك بصيصاً من الأمل في أن تتحلل الطاقة المعتمة في نهاية المطاف (مثل المادة الشبيهة بالطاقة المعتمة التي تم اقتراحها لتفسير توسع الكوزموس)، وإذا حدث ذلك؛ فسيحل **التباطؤ** ***Deceleration*** محل **التعجيل** ***Acceleration***؛ مما يمكّن أشكال الحياة المستقبلية من الاستمرار باستيطان المجرات الجديدة مادامت قائمة.

ما مقدار سرعتك؟

استكشفنا آنفا عدد المجرات التي يمكن لحضارة أن تستوطنها إذا توسعت في كل الاتجاهات بسرعة الضوء. وتقول النسبية العامة إنه من المستحيل إرسال الصواريخ عبر الفضاء بسرعة الضوء؛ لأن هذا سيتطلب طاقة لامنتهية، فما مدى السرعة التي قد تسير بها الصواريخ عمليّاً؟*

حطم صاروخ نيو هورايزن New Horizons rocket التابع لناسا الرقم القياسيّ للسرعة عندما انطلق باتجاه بلوتو في عام 2006 بسرعة نحو 100000 ميل في الساعة (45 كيلومتراً في الثانية)، وفي عام 2018 هدف الصاروخ سولار بروب بلاس Solar Probe Plus -التابع لناسا أيضاً- إلى السفر بأربعة أضعاف هذه السرعة وذلك بالاقتراب من الشمس لاستغلال جاذبيتها كمقلاع، ولكن حتى هذه السرعة هي أقل من 0.1% من سرعة الضوء. و أسر السعيُ إلى الوصول إلى صواريخ أسرع وأفضل خيالَ بعض ألمع العقول في القرن الماضي، وهناك أدب غني ورائع حول هذا الموضوع. لِمَ المضي بسرعة أكبر أمر صعب جدّاً؟ المشكلتان الرئيسيتان هما أن الصواريخ التقليدية تستنفد معظم وقودها

* الرياضيات الكونية بسيطة بشكل مدهش: إذا توسعت الحضارة عبر الفضاء المتوسع ليس بسرعة الضوء c ولكن بسرعة أبطأ v، يقلّ عدد المجرات القابلة للاستيطان بعامل $(v/c)^3$. وهذا يعني أن الحضارات البطيئة تتعرض لعقوبة شديدة، فالحضارة الأبطأ بمعدل عشرة أضعاف، يكون عدد المجرات التي تستوطنها أقل بمقدار 1000 ضعف.

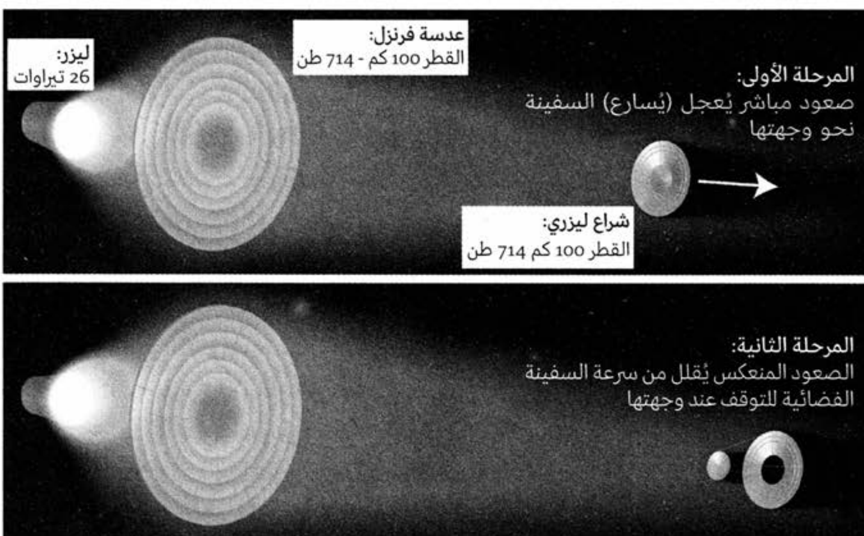
لمجرد تعجيل الوقود الذي تحمله، وأن كفاءة الوقود الصاروخي الحالي منخفضة إلى درجة ميؤوس منها - الجزء من كتلة الوقود الصاروخي الذي يتحول إلى طاقة ليس أفضل بكثير من نسبة 0.0000005% في البنزين التي رأيناها في الجدول 1.6. وأحد التحسينات الواضحة هو التحول إلى وقود أكثر كفاءة. مثلاً، عمل فريمان دايسون وآخرون في مشروع أوريون Orion التابع لناسا، الذي كان يهدف إلى تفجير ما يعادل نحو 300000 قنبلة نووية خلال عشرة أيام للوصول إلى نحو 3%من سرعة الضوء لدفع سفينة فضاء كبيرة بما فيه الكفاية لنقل البشر إلى نظام شمسي آخر في رحلة تمتد إلى قرن من الزمن.[5] وبحث آخرون استخدام ضديد المادة **كوقود**؛ لأن الجمع بينه وبين المادة العادية يطلق الطاقة بنسبة تقارب 100%.

وفكرة أخرى شائعة هي بناء صاروخ لا يحتاج إلى حمل وقوده. مثلاً، الفضاء ما بين النجوم ليس فراغاً تماماً، ولكن يحتوي أحياناً على أيون الهيدروجين (بروتون وحيد: ذرة هيدروجين فقدت إلكترونها). وفي عام 1960 أوحى هذا للفيزيائي روبرت بوسارد Robert Bussard بالفكرة وراء ما يعرف الآن بمُحرك **بوسارد النفَّاث *Bussard ramjet***: لغَرْف مثل هذه الأيونات على الطريق واستخدامها كوقود صواريخ في مفاعل اندماج على متن الصاروخ. وعلى الرغم من أن الأعمال الحديثة أثارت شكوكاً حول إمكان إنجاز هذا العمل في الواقع العملي، فإن هناك فكرة أخرى لا تتضمن حمل الوقود ولكن تبدو ممكنة لحضارة ذات تكنولوجيا فضاء متقدمة: الإبحار بالليزر Laser sailing.

يوضح الشكل 8.6 تصميماً ذكيّاً لصاروخ شراع الليزر الذي ابتكره روبرت فورورد Robert Forward في عام 1984. وهو العالم الفيزيائي نفسه الذي اخترع الأقمار الاصطناعية الثابتة -التي تطرقنا لها عندما تناولنا بناء كرة دايسون. مثلما أن ارتداد جزيئات الهواء عن مركب شراعي يدفعه إلى الأمام، فإن جزيئات الضوء (الفوتونات) التي ترتد من المرآة ستدفع الصاروخ إلى الأمام. وبتسليط إشعاع ليزر ضخم يعمل بالطاقة الشمسية على شراع خفيف لدرجة فائقة متصل بمركبة فضائية، يمكننا استخدام طاقة شمسنا لتسريع الصاروخ إلى سرعات كبيرة، ولكن كيف تتوقف؟ هذا هو السؤال الذي راوغتني إجابته إلى أن قرأت ورقة فورورد الباهرة: كما يظهر في الشكل 8.6، فإن الحلقة الخارجية للشراع الليزر تنفصل وتتحرك أمام المركبة الفضائية، عاكسةً شعاع الليزر إلى الوراء لإبطاء المركبة وشراعها الأصغر.[6] وحسب فورورد الأرقام متوصلاً إلى أن هذا قد يجعل البشر يقطعون رحلة الأربع سنوات الضوئية إلى النظام الشمسي ألفا سنتوري (رجل القنطور) α Centauri في 40 سنة فقط. وبمجرد الوصول إلى هناك، يمكنك أن تتخيل بناء نظام ليزر عملاق جديد ومواصلة التنقل بين النجوم عبر مجرة درب التبانة.

ولكن لماذا التوقف عند هذا الحد؟ في عام 1964. اقترح عالم الفلك السوفييتي نيكولاي كارداشيف Nikolai Kardashev تصنيف الحضارات إلى درجات اعتمادا على مقدار الطاقة التي يمكنها استخدامها. فتسخير طاقة كوكب، أو نجم (باستخدام كرة دايسون، مثلا)، أو مجرة تُشير على التوالي إلى حضارات من النوع الأول Type I والنوع الثاني Type II

والنوع الثالث Type III على مقياس كارداشيف. اقترح مفكرون لاحقون أن النوع الرابع Type IV يجب أن يُشير إلى تسخير طاقة جميع كوننا الذي يمكننا الوصول إليه. منذ ذلك الحين، كانت هناك أخبار سارة وأخبار سيئة لأشكال الحياة الطموحة. الأخبار السيئة هي أن الطاقة المعتمة موجودة، والتي، كما رأينا، تحد من وصولنا. والخبر السار هو التقدم الكبير في الذكاء الاصطناعي. حتى مستشرفو المستقبل المتفائلون مثل كارل ساغان Carl Sagan اعتادوا النظر إلى آفاق وصول البشر إلى المجرات الأخرى على أنها ميؤوس منها إلى حد ما، نظرا إلى نزوعنا إلى أن نموت خلال القرن الأول من رحلة قد تستغرق ملايين السنين حتى لو كانت تسافر بسرعة قريبة من سرعة الضوء. ولما كانوا يرفضون الاستسلام، فقد فكروا في تجميد رواد الفضاء لإطالة حياتهم، وإبطاء شيخوختهم بالسفر بسرعة تقارب سرعة الضوء، أو إرسال مجتمع يسافر لعشرات الآلاف من الأجيال - وهو أطول من تاريخ الجنس البشري بكامله حتى الآن.



(رجل القنطور) α Centauri الواقع على بعد أربع سنوات ضوئية. وفي البداية، يعمل ليزر قوي في نظامنا الشمسي على تعجيل المركبة الفضائية عن طريق تطبيق ضغط الإشعاع على شراع الليزر للمركبة. ولكبح السرعة قبل الوصول إلى الوجهة، ينفصل الجزء الخارجي من الشراع ويعكس ضوء الليزر عن المركبة الفضائية.

إمكانات الذكاء الفائق تغير هذه الصورة تماماً. مما يجعلها واعدةً أكثر بالنسبة إلى أولئك الذين يتشوقون إلى التجوال بين المجرات. أزل الحاجة إلى نقل أنظمة دعم الحياة البشرية الضخمة وأضف تكنولوجيا اخترعها الذكاء الاصطناعي، وفجأة يبدو استيطان ما بين المجرات أمرا سهلا إلى حد ما. يصبح إبحار فورورد باستخدام الليزر أرخص بكثير عندما لا تحتاج المركبة الفضائية إلاّ إلى أن تكون كبيرة بدرجة كافية لاحتواء "مسبار

البذور" Seed probe: روبوت قادر على الهبوط على كويكب أو كوكب في النظام الشمسي المستهدف وبناء حضارة جديدة من المواد الأولية. وليس من الضروري حتى أن يحمل إرشادات: كل ما عليك فعله هو بناء هوائي استقبال كبير بما يكفي لالتقاط مخططات أكثر تفصيلاً وتعليمات مرسلة من الحضارة الأم بسرعة الضوء. بمجرد الانتهاء من ذلك، فإنها ستستخدم أشعة الليزر التي تم إنشاؤها حديثاً لإرسال بعثات بذور جديدة لمواصلة استيطان الأنظمة الشمسية في المجرة واحد بعد الآخر. حتى المساحات المظلمة الشاسعة من الفضاء بين المجرات تميل إلى احتواء عدد كبير من النجوم بين المجرات Intergalactic stars (منبوذة بعد قذفها إلى خارج مجراتها الأصلية) والتي يمكن استخدامها كمحطات طرق؛ مما يعزّز استراتيجية التنقل من جزيرة إلى أخرى للإبحار بالليزر بين المجرات.

بمجرد أن يتم استيطان نظام شمسي أو مجرة أخرى بواسطة الذكاء الاصطناعي الفائق الذكاء، يصبح جلب البشر إلى هناك أمراً سهلاً - إذا نجح البشر في جعل هدفهم هذا هو هدف الذكاء الاصطناعي. يمكن نقل جميع المعلومات الضرورية عن البشر بسرعة الضوء، وبعد ذلك يمكن للذكاء الاصطناعي تجميع الكواركات والإلكترونات في البشر المطلوبين. يمكن القيام بذلك إما بتقنية منخفضة إلى حد ما عن طريق نقل غيغا بايتات من المعلومات اللازمة لتحديد الحمض النووي للشخص ثم حضانة طفل يربيه الذكاء الاصطناعي، أو يمكن للذكاء الاصطناعي أن يبني الكواركات والإلكترونات نانويا في أشخاص مكتملي النمو، يمتلكون جميع الذكريات الممسوحة ضوئيا من أصولهم على الأرض.

هذا يعني أنه إذا كان هناك انفجار ذكاء، فإن السؤال الأساسي ليس ما إذا كان الاستيطان بين المجرات ممكنا، ولكن ببساطة مدى السرعة التي يمكن أن يحدث بها ذلك. لما كانت جميع الأفكار التي ناقشناها آنفا تأتي من بشر، فيجب أن ينظر إليها على أنها الحدود الدنيا على مدى سرعة انتشار الحياة؛ قد تكون حياة فائقة الذكاء طموحة أسرع كثيراً من ذلك، وسيكون لديها حافز قوي لدفع الحدود، لأنه في السباق مع الزمن والطاقة المعتمة، فإن كل زيادة بنسبة 1% في متوسط سرعة الاستيطان تترجم إلى 3% من المجرات المستوطنة.

مثلا، إذا استغرق الأمر 20 عاماً للسفر عشر سنوات ضوئية إلى نظام النجوم التالي باستخدام نظام الشراع بالليزر، ثم عشر سنوات أخرى للاستيطان وبناء أشعة ليزر جديدة ومسابير بذور هناك، فإن المنطقة المستوطنة من الفضاء ستكون مجالاً كرويا ينمو في جميع الاتجاهات بثلث سرعة الضوء في المتوسط. في تحليل جميل وشامل للحضارات المتوسعة عالمياً في عام 2014، نظر الفيزيائي الأمريكي جاي أولسون Jay Olson في بديل عالي التقنية لمقاربة التنقل بين الجزر، تنطوي التقنية على نوعين منفصلين من المسابير: **مسابير البذور والمُوسِّعات *Expanders***.[7] مسابير البذور سوف تتباطأ وتهبط وتبذر بذور الحياة في وجهتها. من ناحية أخرى، لن تتوقف المُوسّعات أبداً: ستغرف المادة أثناء الطيران، وربما

باستخدام بعض التنويعات المحسنة من تكنولوجيا مُحرك بوسارد النفَّاث، وتستخدم هذه المادة كوقود وكمادة أولية تبني منها موسعات ونسخ من نفسها. سيحافظ أسطول المُوسّعات المستنسخ ذاتيا هذا على التعجيل بلطف للحفاظ دائماً على سرعة ثابتة (لنقل نصف سرعة الضوء) نسبة إلى المجرات القريبة، ويتكاثر بوتيرة كافية بحيث ينتشر الأسطول على شكل قشرة كروية متوسعة مع عدد ثابت من المُوسّعات لكل منطقة.

أخيراً وليس آخراً، هناك مقاربة التسلل المراوغ في اللحظة الأخيرة للتوسع بشكل أسرع مما تسمح به أي من الطرق المذكورة آنفا: استخدام عملية "رسائل احتيال كوزموسية" Cosmic spam التي وصفها هانز مورافيك في الفصل الرابع. فعن طريق نشر رسالة تُخدَع الحضارات الساذجة حديثة التطور لبناء آلة فائق الذكاء تختطف حضارتهم، فيمكن للحضارة المخادعة أن تتوسع بسرعة الضوء، بالسرعة التي تنتشر فيها أغنيتها المخاتلة عبر الكون، وقد يكون هذا هو السبيل الوحيد للحضارات المتقدمة للوصول إلى معظم المجرات الواقع ضمن مخروط الضوء المستقبلي الخاص بها، بل ولديهم حافز ضئيل لعدم تجربة ذلك، لذا يجب أن نكون حذرين جدّاً من أي بث من ّخارج الأرض! في كتاب كارل ساغان، اتصال Contact، نحن أبناء الأرض استخدمنا مخططات الكائنات الفضائية لبناء آلة لم نفهمها - أنا لا أوصي بذلك...

باختصار، معظم العلماء والباحثين الذين يفكرون في استيطان الكوزموس، في رأيي، متشائمون تشاؤما مفرطا من حيث تجاهل إمكانات الذكاء الفائق: فبقصر التفكير على المسافرين من البشر، بالغوا في تقدير صعوبة السفر بين المجرات، وبقصر التفكير على تكنولوجيا اخترعها البشر، بالغوا في تقدير الوقت اللازم لمقاربة الحدود المادية لما هو ممكن.

البقاء على اتصال عبر الهندسة الكوزموسية

إذا استمرت الطاقة المعتمة في تعجيل ابتعاد المجرات البعيدة واحدة عن الأخرى، كما تشير أحدث البيانات التجريبية، إذاً سيشكل هذا مصدر إزعاج كبيراً لمستقبل الحياة. فهذا يعني أنه حتى لو تمكنت حضارة مستقبلية من استيطان مليون مجرة، فإن الطاقة المعتمة -عبر مدى من عشرات البلايين من السنين- ستعمل على تشظي هذه الإمبراطورية الكونية إلى آلاف المناطق المختلفة غير القادرة على التواصل مع بعضها البعض. وإذا لم تنجح الحياة في المستقبل في منع هذا التشظي، فستكون أكبر معاقل الحياة المتبقية عبارة عن عناقيد تحتوي على نحو ألف مجرة، والتي تكون جاذبيتها مجتمعة قوياً بدرجة كافية للتغلب على الطاقة المعتمة التي تحاول فصلها عن بعضها.

وإذا رغبت حضارة فائقة الذكاء في البقاء على اتصال، فإن هذا سيعطيها حافزاً قوياً لتطبيق تقنيات هندسة كوزموسية Cosmic engineering واسعة النطاق. ما مقدار المادة التي يتوفر لديها الوقت لنقلها إلى أكبر عنقود فائق Supercluster قبل أن تُبعدها الطاقة المعتمة بعيداً جدّاً عن المنال إلى الأبد؟ إحدى طرق تحريك نجم لمسافات الكبيرة

تتمثل في دفع نجم ثالث نحو نظام ثنائي يدور فيه نجمان بثبات أحدهما حول الآخر. وكما هي الحال في العلاقات الرومانسية، فإن إدخال شريك ثالث قد يزعزع استقرار الأشياء ويؤدي إلى طرد واحد من الثلاثة بعنف - في حالة النجوم، بسرعة كبيرة. إذا كان بعض الشركاء الثلاثة عبارة عن ثقوب سوداء، فيمكن استخدام مثل هذا الثلاثي غير المستقر لقذف كتلة بسرعة كافية لتطير بعيدا عن المجرة المُضيفة. مع الأسف، لا يبدو أن تقنية الأجسام الثلاثة هذه، التي تُطبق إما على النجوم أو الثقوب السوداء أو المجرات، قادرة على تحريك أكثر من جزء صغير من كتلة الحضارة للمسافات الكبيرة المطلوبة للتفوق على الطاقة المعتمة.

لكن من الواضح أن هذا لا يعني أن الحياة الذكية الفائقة لا يمكنها استنباط طرق أفضل، مثلا تحويل الكثير من الكتلة في المجرات البعيدة إلى طاقة وتحميلها على مركبة فضائية يمكنها السفر إلى العنقود الأم. إذا كان من الممكن بناء جهاز سفاليرايز، فيمكن استخدامه حتى لتحويل المادة إلى طاقة تُبثّ كضوء نحو العنقود الأم، إذ يمكن إعادة تنظيمها مرة أخرى كمادة أو استخدامها كمصدر للطاقة.

وسنحالف قمة الحظ إذا اتضح أنه من الممكن بناء ثقوب دودية Wormholes ثابتة قابلة للنقل؛ مما يتيح الاتصال شبه الفوري والسفر بين طرفي الثقب بغض النظر عن المسافة بينهما. والثقب الدودي هو طريق مختصر عبر الزمكان يسمح لك بالسفر من النقطة A إلى النقطة B دون المرور عبر الفضاء المتداخل. وعلى الرغم من أن الثقوب الدودية المستقرة مسموح بها من قبل نظرية آينشتاين للنسبية العامة وقد ظهرت في أفلام مثل اتصال Contact وما بين النجوم Interstellar، فإنها تتطلب وجود نوع افتراضي غريب من المادة ذي كثافة سالبة، والتي قد يتوقف وجودها على تأثيرات الجاذبية الكمية غير المفهومة جيداً. بعبارة أخرى، قد نكتشف أن الثقوب الدودية المفيدة مستحيلة. لكن إن لم تكن كذلك، فستكون لدى ذكاء فائق حوافز ضخمة لبناء مثل هذه الثقوب. لن تحدث الثقوب الدودية ثورة في التواصل السريع داخل المجرات فحسب، بل وأيضا بربط المجرات على الأطراف البعيدة بالعنقود المركزي في وقت مبكر، أي ستسمح الثقوب الدودية لجميع مناطق الحياة المستقبلية في البقاء على تواصل مع بعضها البعض عبر المسافات البعيدة، مُعرقلة تماماً محاولات الطاقة المعتمة لإحباط الاتصالات. وبمجرد أن ترتبط مجرتان من خلال ثقب دودي مستقر، فإنهما ستبقيان متصلتين بغض النظر عن المسافة التي تنجرفان فيها مبتعدتين إحداهما عن الأخرى.

إذا وجدت حضارة مستقبلية -وعلى الرغم من أفضل محاولاتها في الهندسة الكوزموسية- أن أجزاء منها محكوم عليها بالانجراف وفقدان الاتصال إلى الأبد، فقد تسمح لها ببساطة بالرحيل وتتمنى لها التوفيق. ولكن، إذا كان لديها أهداف حوسبية طموحة تتضمن البحث عن إجابات لبعض الأسئلة الصعبة جدّاً، فقد تلجأ بدلاً من ذلك إلى استراتيجية القطع والحرق: يمكنها تحويل المجرات البعيدة إلى أجهزة حاسوب ضخمة تحول مادتها وطاقتها إلى حوسبات بسرعات مسعورة، على أمل أنه قبل أن تدفع

الطاقةُ المعتمة بقاياها المحترقة بعيدا عن الأنظار، سيكون بإمكانها بث الإجابات التي طال انتظارها إلى العنقود الأم. ستكون استراتيجية القطع والحرق هذه مناسبة بشكل خاص للمناطق البعيدة التي لا يمكن الوصول إليها إلا من خلال منهجية "رسائل الاحتيال الكوزموسية"، على الرغم من امتعاض السكان الموجودين في الأصل. أما في الموطن، في المنطقة الأم، فيمكن للحضارة بدلاً من ذلك أن تهدف إلى تحقيق أقصى درجات المحافظة والكفاءة لأطول فترة ممكنة.

إلى متى يمكنك الصمود؟

طول العمر هو شيء يطمح إليه معظم الأشخاص والمنظمات والدول. لذلك إذا طورت حضارة مستقبلية طموحة ذكاءً فائقا وتريد طول العمر، فإلى متى يمكنها الصمود؟

أول تحليل علمي دقيق لمستقبلنا البعيد أجراه فريمان دايسون لا غيره. ويلخص الجدول 6.3 بعضا من أهم استنتاجاته الرئيسية. والخلاصة هي أنه ما لم يتدخل الذكاء، فإن الأنظمة الشمسية والمجرات تُدمَّر تدريجيّاً، ويتبعها في النهاية كل شيء آخر، ولا تترك سوى فضاء بارد ميت وفارغ بتوهج آخذ في الخفوت إلى الأبد. لكن فريمان ينهي تحليله بملاحظة متفائلة: «هناك أسباب علمية جيدة لأخذ احتمال أن تنجح الحياة والذكاء في تشكيل كوننا هذا لأغراضهما الخاصة على محمل الجد».[8]

ماذا	متى
العصر الحالي لكوننا	10^{10} سنة
الطاقة المعتمة تدفع معظم المجرات بعيدا عن المنال	10^{11} سنة
آخر النجوم تحترق	10^{14} سنة
الكواكب تنفصل عن النجوم	10^{15} سنة
النجوم تنفصل عن المجرات	10^{19} سنة
اضمحلال المدارات بفعل إشعاع الجاذبية	10^{20} سنة
اضمحلال البروتونات (في أقرب وقت ممكن)	$> 10^{34}$ سنة
تبخر الثقوب السوداء ذات الكتلة النجمية	10^{67} سنة
تبخر الثقوب السوداء الهائلة	10^{91} سنة
يتحلّل جميع المادة إلى حديد	10^{1500} سنة
جميع المواد تشكل ثقوباً سوداء تتبخر بعد ذلك	$10^{10^{26}}$ سنة

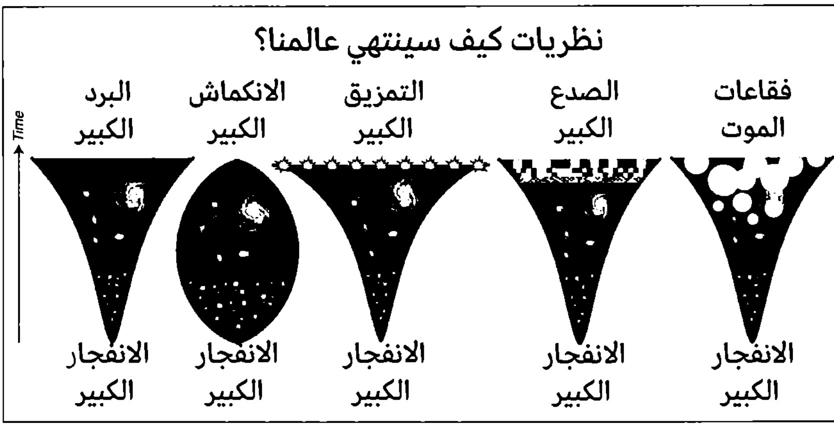
الجدول 3.6: تقديرات المستقبل البعيد، كل ما عدا التقدير الثاني والسابع هي تقديرات من إعداد فريمان دايسون/ لقد أجرى هذه الحسابات قبل اكتشاف الطاقة المعتمة، التي قد تمكن عدة أنواع من "انهيار الكوزموس" Cosmocalypse في 10^{10} - 10^{11} سنة. قد تكون البروتونات مستقرة تماماً؛ لكن إذا لم يكن الأمر كذلك، تشير التجارب إلى أن الأمر سيستغرق أكثر من 10^{34} عام حتى يتحلل نصفها.

أعتقد أن الذكاء الفائق يمكن أن يحل الكثير من المشكلات المذكورة في الجدول 3.6. إذ يمكن إعادة ترتيب المادة إلى شيء أفضل من الأنظمة الشمسية والمجرات. والتحديات التي كثيرا ما تُناقش، مثل موت شمسنا خلال بضعة بلايين من السنين، لن تشكل عائقا، لأنه حتى حضارة ذات تكنولوجيا منخفضة نسبياً يمكن أن تنتقل بسهولة إلى نجوم منخفضة الكتلة تدوم لأكثر من 200 بليون سنة. على افتراض أن حضارات فائقة الذكاء ستبني محطاتها لتوليد الطاقة التي هي أكثر فعالية من النجوم، فإنها قد ترغب في الواقع في منع تكوين النجوم للحفاظ على الطاقة: حتى لو استخدموا كرة دايسون لحصاد جميع مخرجات الطاقة خلال عمر النجم الرئيسي (استخلاص نحو 0.1% من إجمالي الطاقة)، فإنها قد تكون غير قادرة على الحيلولة دون هدر جزء كبير من 99.9% المتبقية من الطاقة عندما تموت النجوم الضخمة جدا. يموت نجم ثقيل في انفجار سوبرنوفا (مستعر أعظم) Supernova تنطلق منه معظم الطاقة كنيوترينوات مراوغة، وبالنسبة إلى النجوم الثقيلة جدّاً، تضيع كمية كبيرة من الكتلة من خلال تشكيل ثقب أسود تتسرب منه الطاقة شيئا فشيئا على مدى 10^{67} سنة.

مادامت الحياة فائقة الذكاء لم تنفد من المادة/الطاقة، فيمكنها الحفاظ على موئلها Habitat في المكان الذي ترغب فيه.وقد تكتشف هذه الحياة طريقة لمنع تحلل البروتونات باستخدام ما يسمى بتأثير الوعاء المرصود Watched-pot effect في ميكانيكا الكم، حيث تتباطأ عملية التحلل بسبب الرصد المنتظم. ولكن هناك معيق محتمل: انهيار الكوزموس مدمرا كوننا بأجمعه، وفي وقت قريب نحو 10 - 100 بليون سنة من الآن. فاكتشاف الطاقة المعتمة والتقدم في نظرية الأوتار String theory أثار سيناريوهات جديدة حول انهيار الكون لم يكن فريمان دايسون على دراية بها عندما كتب ورقته المؤثرة.

إذاً كيف سينتهي عالمنا، بعد بلايين السنين من الآن؟ لديّ خمسة مشتبه منهم رئيسيون في كيفية انهيار العالَم Apocalypse، أو انهيار الكوزموس Cosmocalypse، موضحة في الشكل 9.6: **البرد الكبير *Big Chill***، **والانكماش الكبير *Big Crunch***، **والتمزّق الكبير *Big Rip***، **والصدع الكبير *Big Snap***، **وفقاعات الموت *Death Bubbles***، فعالمنا الحالي يتوسع منذ نحو 14 بليون سنة. وسيحل البرد الكبير هو عندما يستمر الكون في التوسع إلى الأبد، مخفِّفا كوننا فيتحول إلى مكان بارد ومظلم وميت في النهاية، وقد كان ينظر إلى هذه النهاية على أنها النهاية الأكثر ترجيحاً عندما كتب فريمان تلك الورقة البحثية. وأنا أرى أن ذلك يشبه خيار الشاعر تي. إس. إيليوت T. S. Eliot: «هذه هي الطريقة التي ينتهي بها العالم / ليس بضجة ولكن بأنين». إذا كنت، مثل روبرت فروست Robert Frost، تفضل أن ينتهي العالم بحريق بدلاً من الجليد، إذاً تمنَّ حدوث الانكماش الكبير، عندما ينعكس التوسّع الكوني في نهاية المطاف وينهار كل شيء انهيارا كارثيا يشبه الانفجار الكبير ولكن معكوسا. وأخيرا، فإن التمزق الكبير قد يبدو مثل البرد الكبير لفارغ الصبر، إذ تتمزق مجراتنا وكواكبنا وحتى ذراتنا في ختام نهائي بعد وقت محدد من الآن. أي من هذه النظريات الثلاث يجب أن تراهن عليها؟ يعتمد ذلك على ما ستفعله الطاقة المعتمة،

التي تشكل نحو 70% من كتلة عالمنا. مع استمرار التوسع في الفضاء. وقد يقع أي من سيناريوهات البرد أو السحق أو التمزيق، اعتماداً على ما إذا كانت الطاقة المعتمة تبقى موجودة دون تغيير أو تخفف وصولا إلى كثافة سلبية أو تتراكم إلى الكثافة الأعلى، على التوالي. ونظراً لأننا لا نفهم الطاقة المعتمة، سأخبرك أين أضع رهاني: 40% على البرد الكبير، 9% على الانكماش الكبير، 1% على التمزيق الكبير.



الشكل 9.6: نعلم أن عالمنا بدأ بانفجار كبير ساخن قبل 14 بليون عام، وتوسّع وبرد، ودمج جسيماته في شكل ذرات ونجوم ومجرات. لكننا لا نعرف مصيره النهائي. السيناريوهات المقترحة البرد الكبير (التوسع السرمدي)، الانكماش الكبير (الانهيار)، التمزيق الكبير (معدل تسارع لا منته يمزق كل شيء)، الصدع الكبير (نسيج الفضاء يتكشف عن طبيعة مُحبَّبة Granular عند تمدده تمدداً شديداً)، فقاعات الموت ("تجمد" الفضاء في فقاعات مميتة تتسع بسرعة الضوء).

ماذا عن 50% الأخرى من رهاني؟ أحتفظ بذلك لخيار "لا شيء مما سبق"، لأنني أعتقد أننا البشر يجب أن نكون متواضعين ونعترف بأن هناك أشياء أساسية ما زلنا لا نفهمها. طبيعة الفضاء، مثلاً: تفترض نهايات البرد والسحق والصدع أن الفضاء نفسه مستقر وقابل للامتداد إلى ما لا نهاية. فقد اعتدنا أن نفكر في الفضاء باعتباره مجرد مسرح ثابت وممل تتكشف على منصته الدراما الكونية. ثم علمنا آينشتاين أن الفضاء هو بالفعل أحد العناصر الفاعلة الرئيسية: يمكن أن ينحني إلى ثقوب سوداء، ويمكن أن يتحول إلى موجات جاذبية (ثقالية) Gravitational waves ويمكن أن يتمطّط ككون ممتد. وربما قد يتجمد مثل الماء في مرحلة مختلفة تماماً. و في هذه المرحلة الجديدة تُشكِّل فقاعات الموت السريعة التوسع مرشحاً آخر لانهيار الكون. إذا كانت فقاعات الموت محتملة، فمن المرجح أنها ستتوسع بسرعة الضوء، تماماً مثل المجال المتنامي لرسائل الاحتيال الكوزموسية من حضارة شديدة العدوانية.

إضافة إلى ذلك، تقول نظرية آينشتاين إن تمدّد الفضاء قد يستمر إلى الأبد؛ مما يسمح للكون بالاقتراب من حجم لا منتهٍ كما هي الحال في سيناريوهات التجمد الكبير

والصدع الكبير. هذا يبدو جيداً لدرجة تبعث على الشك في صحته، وأنا أظن ذلك. فالشريط المطاطي يبدو سلسا ومستمراً، تماماً مثل الفضاء، ولكن إذا مططته أكثر من اللازم، فإنه ينقطع، لماذا ؟ نظراً لأنه مصنوع من الذرات، وبتمطيط كافٍ، تصبح هذه الطبيعة الذرية المُحبّبة Granular atomic nature للمطاط مهمة. فهل من الممكن أن يكون للفضاء نوعاً من الطبيعة المُحبّبة ولكن على مستوى صغير جدّاً لدرجة أننا لا نرصدها؟ تشير أبحاث الجاذبية الكمية إلى أنه من غير المنطقي الحديث عن الفضاء ثلاثي الأبعاد التقليدي على نطاقات أصغر من نحو 10^{-34} متر. إذا كان صحيحاً حقاً، لا يمكن مطّ الفضاء إلى ما لا نهاية دون انهيار الكون في حدثِ صدعٍ كبير، إذاً قد ترغب الحضارات المستقبلية في الانتقال إلى أكبر منطقة في الفضاء لا تتمدد (عنقود مجري ضخم) يمكنهم الوصول إليها.

ما المقدار الذي يمكنك حوسبته؟

بعد استكشاف المدة التي **يمكن** أن تدومها الحياة في المستقبل، دعنا نستكشف المدة التي قد **ترغب** الحياة في أن تستمر في الوجود طوالها. على الرغم من أنه قد يكون من الطبيعي أن ترغب في العيش لأطول فترة ممكنة، فقد قدم فريمان دايسون حجة كميّة Quantitative argument لهذه الرغبة: تنخفض تكلفة الحوسبة عندما تُجرى الحوسبة ببطء، لذلك ستحصل في النهاية على المزيد إذا بطّأت الأمور إلى أدنى قدر ممكن. ووفقا لفريمان فأنه إذا استمر كوننا في التوسع والتبريد إلى الأبد، فربما يكون هناك قدر لا منتهٍ من الحوسبة.

البطء لا يعني الملل بالضرورة: إذا كانت الحياة المستقبلية تعيش في عالم مُحاكى، فليس للزمن الذي تخبره ذاتيا أي علاقة بالوتيرة الجليدية التي تجري بها المحاكاة في العالم الخارجي، ومن ثم فإن احتمالات حوسبة لا *منتيهة* يمكن أن تترجم إلى خلود ذاتي لأشكال الحياة المُحاكاة. اعتمد عالم الكوزمولوجيا فرانك تبلر Frank Tipler على هذه الفكرة للتكهن بأنه يمكنك أيضاً تحقيق الخلود الذاتي Subjective immortality في اللحظات الأخيرة قبل الانكماش الكبير من خلال تسريع العمليات الحوسبية نحو اللانهاية مع الارتفاع الشاهق لدرجة الحرارة والكثافة.

ولما كان من الظاهر أن الطاقة المعتمة تُفسد كلاً من أحلام فريمان وفرانك في الحوسبة اللامنتيهة، فقد يُفضّل الذكاء الفائق في المستقبل أن يحترق بإمداداته من الطاقة بسرعة نسبية، ليحوّلها إلى حوسبات قبل أن يواجه مشكلات مثل الآفاق الكونية واضمحلال البروتون. إذا كان تعظيم الحوسبة هو الهدف الأقصى، فستكون أفضل استراتيجية هي المقايضة بين أبطأ من اللازم (لتجنب المشكلات المذكورة آنفا) وأسرع من اللازم (منفقة طاقة أكثر مما هو مطلوب في كل عملية حوسبية).

وإذا جمعنا كل شيء استكشفناه في هذا الفصل، فإنه سيخبرنا أن محطات الطاقة والحواسيب الفعالة فعالية قصوى ستمكن حياة فائقة الذكاء من أداء قدر الحوسبة تُحيّر

العقل. فتشغيل دماغك ذي 13 وات لمئة عام يتطلب مقدار الطاقة الموجود في نحو نصف مليغرام من المادة - أقل مما هو موجود في حبة سكر. ويقترح عمل سيث لويد Seth Lloyd أنه من الممكن جعل الدماغ أكثر كفاءة من حيث الطاقة بمقدار كوادرليون ضعف؛ مما يمكّن حبة السكر تلك من تشغيل محاكاة جميع الحيوات البشرية التي وجدت إضافة إلى آلاف أضعاف ذلك العدد من البشر. إذا كان بالإمكان استخدام جميع المادة الموجودة في عالمنا المتاح لمحاكاة أشخاص، فإن ذلك سيمكن 10^{69} حياة - أو أي شيء آخر يفضل الذكاء الاصطناعي الفائق القيام به بقوته الحوسبية. وسيكون من الممكن تحقيق المزيد من الحيوات إذا تم تشغيل عمليات المحاكاة الخاصة بها بسرعة أبطأ.[9] على العكس من ذلك، قدّر نيك بوستروم في كتابه الذكاء الفائق Superintelligence، أنه يمكن محاكاة 10^{58} حياة بشرية بافتراضات أكثر تحفظاً فيما يختص بكفاءة الطاقة. ولكننا نُجزِّئ هذه الأرقام لتحليلها، فهي ضخمة، وكذلك هي مسؤوليتنا في ضمان عدم تبديد هذه الإمكانات المستقبلية للحياة. كما يقول بوستروم: «إذا كنا نمثل جميع السعادة التي عشناها في حياة واحدة بكاملها بدمعة فرح واحدة، فإن سعادة هذه النفوس ستملأ وستعيد ملء محيطات الأرض كل ثانية، وتواصل القيام بذلك لمئة بليون بليون ألفية من السنين، من المهم حقاً أن نتأكد من أن هذه الدموع دموع فرح حقا».

تسلسلات هرمية كوزموسية

لا تحد سرعة الضوء من انتشار الحياة فحسب، بل تحدِّد أيضاً طبيعة الحياة، وتضع قيوداً شديدة على التواصل والوعي والسيطرة. لذلك إذا انتشرت الحياة في الكثير من كوننا في نهاية المطاف، فكيف ستكون هذه الحياة؟

تسلسلات هرمية فكرية

هل سبق لك أن حاولت سحق ذبابة بيدك وفشلت؟ السبب هو أن الذبابة تتفاعل أسرع منك لأنها أصغر حجماً، بحيث يستغرق نقل المعلومات بين العينين والدماغ والعضلات وقتاً أقل. ومبدأ «أكبر = أبطأ» لا ينطبق فقط على البيولوجيا، حيث يتحدَّد الحد الأقصى للسرعة بالسرعة التي يمكن أن تنتقل بها الإشارات الكهربائية عبر العصبونات (الخلايا العصبية) Neurons، ولكن أيضاً على الحياة الكوزموسية المستقبلية إذا لم يكن بإمكان المعلومات الانتقال بأسرع من الضوء. لذلك، فلنظام معالجة معلومات ذكي، فإن الكِبر في الحجم هو نعمة مختلطة تنطوي على مقايضة مثيرة للاهتمام. من ناحية، يتيح لك كبر الحجم يبطىء القدرة على احتواء المزيد من الجسيمات، التي تمكن لأفكار أكثر تعقيداً. من ناحية أخرى، يبطىء هذا المعدلَ الذي يمكن به لهذا الحجم أن يكون لديه أفكار عالمية حقيقية، لأن نشر المعلومات ذات الصلة إلى جميع أجزائه يستغرق وقتاً أطول.

إذاً إذا غمرت الحياة كوزموسنا، فما الشكل الذي ستختاره: بسيط وسريع، أو معقد وبطيء؟ أتوقع أنها ستتخذ نفس خيار الحياة على الأرض: كلاهما! فبيوسفير (المحيط الحيوي) الأرض يمتد عبر نطاق مذهل من الأحجام، من الحيتان الزرقاء العملاقة التي يبلغ وزنها مئتي طن نزولا إلى البكتيريا الدقيقة **بيليغيباكتر** ***Pelagibacter*** التي تزن نحو 10^{-16} كيلوغرام، والتي يُعتقد أن مجموعها يمثل كتلة حيوية أكثر من جميع أسماك العالم مجتمعة. إضافة إلى ذلك، فإن الكائنات الكبيرة والمعقدة والبطيئة غالباً ما تخفف من آثار بطئها باحتوائها على وحدات أصغر بسيطة وسريعة. مثلاً، رد فعل الرمش بالعين هو سريع جدّاً بالذات لأنه يُنفذ بواسطة دائرة صغيرة وبسيطة لا تشمل معظم عقلك: إذا اتجهت تلك الذبابة التي يصعب سحقها باليد خطأ نحو عينيك، فسترمش في غضون عشر ثانية، قبل وقت طويل من أن يُتاح للمعلومات ذات الصلة أن تنتشر في جميع أنحاء عقلك وتجعلك تدرك بوعي ما حدث. وبتنظيم معالجة المعلوماته في تسلسل هرمي من الوحدات، يتمكن بيوسفيرنا من الاحتفاظ بالكعكة وأكلها، مُحقِّقا السرعة والتعقيد. ونحن البشر نستخدم بالفعل الاستراتيجية الهرمية نفسها لتحسين الحوسبة المتوازية Parallel computing.

نظراً لأن الاتصال الداخلي بطيء ومكلف، أتوقع أن تفعل الحياة الكوزموسية المستقبلية المتقدمة الشيء نفسه، بحيث تُجرى الحوسبات على المستوى المحلي قدر الإمكان. إذا كانت عمليةً حوسبية بسيطة بما يكفي ليقوم بها حاسوب بحجم كيلوغرام واحد، فمن غير المفيد حوسبتها على حاسوب بحجم المجرة، نظراً لان انتظار وصول المعلومات المرسلة بسرعة الضوء بعد كل خطوة حوسبية سيسبب تأخيرا طويلا يعادل نحو 100,000 سنة لكل خطوة.

ما إذا، إن وجدت أصلا، كانت عملية معالجة المعلومات المستقبلية هذه **واعية** Conscious بمعنى أن لها تجربة ذاتية، هو موضوع مثير للجدل ورائع سنستكشفه في الفصل الثامن. إذا كان الوعي يتطلب أن تكون أجزاء مختلفة من النظام قادرة على التواصل مع بعضها البعض، إذاً فإن أفكار المنظومات الأكبر بطيئة بالضرورة. وبينما يمكن أن تكون لديك أنت -أو حاسوب عملاق بحجم الأرض في المستقبل- العديد من الأفكار في الثانية الواحدة، فقد يكون لعقل بحجم المجرة فكرة واحدة فقط كل مئة ألف عام، وعقل كوزموسي يبلغ حجمه بليون سنة ضوئية لن يكون لديه سوى وقت لنحو عشر أفكار في المجمل قبل أن يتشظى إلى أجزاء منفصلة بفعل الطاقة المعتمة. من ناحية أخرى، قد تكون هذه الأفكار الثمينة القليلة والخبرات المصاحبة عميقة جدا!

تسلسلات هرمية سلطوية

إذا تم تنظيم الفكر بحد ذاته في مجموعة من التسلسلات الهرمية التي تمتد عبر مجموعة واسعة من المقاييس، فماذا عن السلطة؟ في الفصل الرابع. استكشفنا الكيفية

اللتي تنظم بها الكيانات الذكية نفسها بشكل طبيعي في تسلسل هرمي للسلطة في توازن ناش، حيث سيكون أي كيان أسوأ حالاً إذا غير استراتيجيته. وإذا صارت تكنولوجيا الاتصالات والنقل أفضل، زاد حجم هذه التسلسلات الهرمية. إذا توسع الذكاء الفائق إلى أحجام فلكية، فما شكل هرم قوتها؟ هل ستكون حرة ولامركزية أم استبدادية جدا؟ هل سيعتمد التعاون بشكل أساسي على المنفعة المتبادلة أم الإكراه والتهديدات؟

لتسليط الضوء على هذه الأسئلة، دعونا نأخذ بعين الاعتبار الجزرة والعصا (الحوافز والتهديدات): ما هي الحوافز الموجودة لتعزيز التعاون على النطاقات الكونية، وما هي التهديدات التي يمكن استخدامها لفرض التعاون؟

السيطرة بالجزرة

على الأرض، كانت **التجارة** محركاً تقليدياً للتعاون نظرا إلى أن إنتاج الأشياء عبر الكوكب يتفاوت نسبيا من حيث الصعوبة. إذا كانت تكلفة تعدين كيلوغرام واحد من الفضة أكثر بنحو 300 ضعف من تعدين كيلوغرام واحد من النحاس في منطقة ما، ولكن أكثر بنحو 100 ضعف في منطقة أخرى، فسيقايض كل منهما 200 كيلوغرام من النحاس مقابل كيلوغرام من الفضة. وإذا كان لدى إحدى المناطق تقنية أكثر تقدما من منطقة أخرى، فبإمكان كليهما الاستفادة بشكل مماثل من تجارة السلع عالية التقنية مقابل المواد الخام.

ولكن، إذا طور ذكاء فائق تكنولوجيا يمكنها إعادة ترتيب الجسيمات الأولية بسهولة إلى أي شكل من أشكال المادة، فسوف تقضي على معظم حوافز التجارة عبر المسافات الطويلة. لماذا تهتم بشحن الفضة بين الأنظمة الشمسية البعيدة عندما يكون تحويل النحاس إلى الفضة أبسط وأسرع من خلال إعادة ترتيب جزيئاته؟ لماذا تهتم بشحن الآلات عالية التقنية بين المجرات عندما تكون كل من الدراية والمواد الخام (يمكن استخدام أي مادة) موجودة في كلا المكانين؟ أظن أنه في كونٍ يعج بالذكاء الفائق، فإن السلعة الوحيدة التي تستحق الشحن لمسافة طويلة هي المعلومات Information. الاستثناء الوحيد قد يكون المادة التي ستستخدم لمشاريع الهندسة الكونية - مثلاً، التصدي للاتجاه المدمر -المذكور آنفا- للطاقة المعتمة التي تمزق الحضارات. على عكس التجارة البشرية التقليدية، يمكن شحن هذه المادة بأي كمية من أين مكان، ربما حتى كحزمة طاقة، لأن الذكاء الفائق المتلقي يمكن أن يعيد ترتيبها بسرعة إلى أي شيء يريده.

إذا برزت المشاركة أو تبادل المعلومات كمحرك رئيسي للتعاون الكوني، فعندئذ ما نوع المعلومات المتضمنة في تلك التعاملات؟ أي معلومات مرغوبة ستكون ذات قيمة إذا كان إنشاؤها يتطلب جهدا حوسبيا هائلا ويستغرق وقتا طويلا. مثلاً، قد يرغب ذكاء فائق في الحصول على إجابات للأسئلة العلمية الصعبة حول طبيعة الواقع المادي، والمسائل الرياضياتية الصعبة للنظريات والخوارزميات المثلى، والأسئلة الهندسية الصعبة حول أفضل طريقة لبناء تكنولوجيا مدهشة. قد تحتاج أشكال الحياة التلذّذيّة (نسبة إلى المذهب التلذّذي

Hedonistic) إلى ترفيه رقمي وتجارب محاكاة رائعة، وقد تُشعل التجارة الكونية زيادة الطلب على بعض أشكال العملة الكونية المُرمَّزة Cryptocurrency من مثل البيتكوين.

قد تحفز فرص تقاسم المعلومات هذه، ليس فقط الكيانات ذات القوة المتساوية تقريباً، ولكن أيضاً التسلسلات الهرمية للسلطة صعوداً وهبوطاً، مثلا بين عُقد Nodes بحجم النظام الشمسي والمراكز المجرِّية أو بين عُقد مجرّية ومراكز كونية، قد ترغب العُقد في ذلك لمجرد أن تكون جزءاً من شيء أكبر، ولتزويدها بالإجابات والتقنيات التي لا يمكنها تطويرها بنفسها، والدفاع ضد التهديدات الخارجية. وقد يرغبون أيضاً بوعد شبه الخلود من خلال النسخ الاحتياطي Backups: تماماً مثلما يجد كثير من البشر العزاء في اعتقادهم أن عقولهم ستعيش بعد موت أجسادهم المادية. قد يُقدِّر الذكاء الاصطناعي المتقدم أن عقله ومعرفته ستستمر في مركز حاسوب فائق بعد أن تستنفد أجهزته المادية الأصلية احتياطياتها من الطاقة.

على العكس من ذلك، قد يرغب المركز في أن تساعده عُقده على أداء مهام حوسبة ضخمة على المدى الطويل عندما لا تكون هناك حاجة فورية إلى النتائج، بحيث يستحق الأمر انتظار آلاف أو ملايين السنين للحصول على الإجابات. كما استكشفنا آنفا، قد يرغب المركز أيضاً في أن تساعده عُقده على تنفيذ مشاريع هندسية كونية ضخمة مثل عكس الطاقة المعتمة المدمرة عن طريق تحريك تركيزات الكتلة المجرية. إذا اتضح أن الثقوب الدودية -التي تسمح بالانتقال عبرها- ممكنة وقابلة للبناء، فقد تكون الأولوية العليا للمركز هي بناء شبكة من هذه الثقوب الدودية لإحباط الطاقة المعتمة وإبقاء إمبراطوريتها متصلة بشكل دائم. إن الأسئلة المتعلقة بالأهداف النهائية لذكاء فائق كوني هي أسئلة رائعة ومثيرة للجدل سنستكشفها أكثر في الفصل السابع.

السيطرة بالعصا

عادة ما تجبر إمبراطوريات الأرض الخاضعين لها على التعاون باستخدام الجزرة والعصا. بينما قدّر رعايا الإمبراطورية الرومانية التكنولوجيا والبنية التحتية والدفاع الموفَّرة لهم كمكافأة على تعاونهم، غير أنهم كانوا يخشون أيضاً تداعيات التمرد أو عدم دفع الضرائب. وبسبب الوقت الطويل الذي يتطلبه إرسال قوات من روما إلى المقاطعات النائية، تم تفويض جزء من التخويف إلى قوات محلية وموظفين موالين لهم صلاحية فرض عقوبات شبه فورية. ويمكن لمركز ذكاء فائق أن يستخدم استراتيجية شبيهة متمثلة في نشر شبكة من الحراس الموالين في جميع أنحاء الإمبراطورية الكوزموسية، ونظراً لصعوبة التحكم في رعايا ذكاء فائق، فقد تتمثل أبسط الاستراتيجيات القابلة للتطبيق في استخدام حراس ذكاء اصطناعي مبرمجين ليكونوا مخلصين بنسبة 100% وبسيطين نسبياً، فقط يراقبون ما إذا كانت جميع القوانين مطبقة، وتلقائياً يشغّلون جهاز يوم الهلاك بخلاف ذلك.

لنفترض، مثلاً، أن مركز الذكاء الاصطناعي يُعدّ العُدة لوضع قزم أبيض بالقرب من حضارة بحجم النظام الشمسي ترغب في السيطرة عليها. وقزم أبيض هو عبارة عن قشرة

محترقة لنجم ثقيل نسبيا، تتألف إلى حد كبير من الكربون، فهي تشبه ألماسة عملاقة في السماء، وهي مضغوطة لدرجة يمكن أن يكون وزنها أكثر من الشمس في حين تكون أصغر من الأرض حجما. أثبت عالِم الفيزياء الهندي سابرامانين تشاندراسيخار Subrahmanyan Chandrasekhar أنه في حالة استمرارك في إضافة كتلته إلى أن تتجاوز **حد تشاندراسيخار** ***Chandrasekhar limit***، أي نحو 1.4 ضعف كتلة شمسنا، فستنفجر انفجارا نوويا حراريا كارثيا يُعرف بالسوبرنوفا (المستعرات العظمى) Supernova من النوع 1A. وإذا كان مركز الذكاء الاصطناعي قد رتب بقسوة لوضع هذا القزم الأبيض قريباً جدّاً من حد تشاندراسيخار، فإن الحارس الذكاء الاصطناعي قد يكون فعالاً حتى لو كان بسيطاً جدّاً (في الواقع، إلى حد كبير لأنه كان بسيطا جدّاً): يمكن برمجته للتحقق ببساطة من أن الحضارة الخاضعة قد قد دفعت حصتها الشهرية من عملات البيتكوين الكونية، أو البراهين الرياضياتية أو أي ضرائب أخرى منصوص عليها، وإذا لم يكن الأمر كذلك، يلقي الذكاء الاصطناعي بكتلة كافية على القزم الأبيض لإشعال السوبرنوفا وتفجير المنطقة بأسرها إلى فتات.

وبالمثل يمكن السيطرة على حضارات بحجم مجرة بوضع أعداد كبيرة من الإجسام الكثيفة في مدارات قريبة حول ثقب أسود ضخم في مركز المجرة، والتهديد بتحويل هذه الكتل إلى غاز، مثلا عن طريق مصادمتها ببعضها، عندئذ سيبدأ هذا الغاز بتغذية الثقب الأسود، ويحوله إلى كوازار قوي؛ مما يجعل الكثير من المجرة غير صالح للسكنى.

باختصار، هناك حوافز قوية للحياة المستقبلية للتعاون مع بعضها عبر مسافات فلكية. لكن السؤال المطروح بشكل واسع يدور حول ما إذا كان هذا التعاون سيعتمد بشكل أساسي على المنافع المتبادلة أو على التهديدات الشرسة -يبدو أن الحدود التي فرضتها الفيزياء تسمح بكلا السيناريوهين، ومن ثم فإن النتيجة تعتمد على الأهداف والقيم السائدة. وسنستكشف قدرتنا على التأثير في أهداف وقيم الحياة المستقبلية في الفصل السابع.

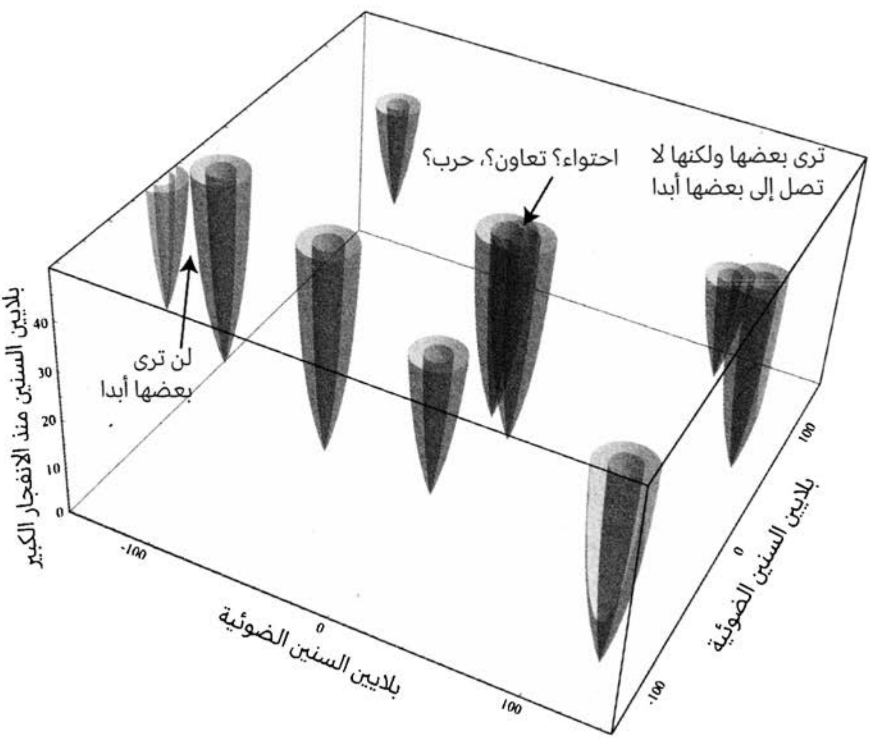
عندما تتصادم الحضارات

حتى الآن، ناقشنا فقط السيناريوهات التي انتشرت فيها الحياة عبر كوننا من انفجار ذكاء واحد. ولكن ما الذي سيحدث إذا تطورت الحياة بشكل مستقل في أكثر من مكان والتقت حضارتان متوسعتان؟

إذا أخذت بعين الاعتبار عشوائيا نظاما شمسيا، فهناك قدر من الاحتمال بأن تتطور الحياة على أحد كواكبها، مطوّرة تكنولوجيا متقدمة والتوسع في الفضاء. ويبدو أن هذا الاحتمال أكبر من الصفر نظرا لأن الحياة التكنولوجية تطورت هنا في نظامنا الشمسي، ويبدو أن قوانين الفيزياء تسمح باستيطان الفضاء. إذا كان الفضاء كبيرا بما فيه الكفاية (في الواقع، تشير نظرية التوسع الكوني إلى أنه واسع أو لامنتهٍ)، فمن ثم سيكون هناك العديد من الحضارات المتوسعة، كما هو موضح في الشكل 10.6. وتتضمن ورقة جاي أولسون البحثية المذكورة آنفا تحليلاً أنيقاً لمثل هذه البيوسفيرات الكونية المتوسعة،

وأجرى توبي Toby Ord أورد تحليلاً مشابهاً مع زملائه في معهد مستقبل الإنسانية Future of Humanity Institute. إذا نظرنا إليها في ثلاثة أبعاد، فإن هذه البيوسفيرات الكونية هي حرفيا عبارة عن كرات Spheres مادامت الحضارات تتوسع بالسرعة نفسها في جميع الاتجاهات. وفي الزمكان، تشبه الجزء العلوي من الكأس الزجاجية في الصورة 10.6، لأن الطاقة المعتمة تحد من عدد المجرات التي يمكن أن تصل إليها كل حضارة.

إذا كانت المسافة بين حضارات متجاورة من الحضارات المُستوطنة للفضاء أكبر بكثير مما تسمح به الطاقة المعتمة لها بالانتشار، فلن تتصل أبدا ببعضها البعض أو حتى تعرف بوجود بعضها البعض. لذلك سيشعرون كما لو كانوا وحدهم في الكوزموس. وإذا كان كوزموسنا أكثر خصوبة، بحيث يكون الجيران أقرب إلى بعضهم البعض، فإن بعض الحضارات ستتقاطع في النهاية. ماذا سيحدث في هذه المناطق المتقاطعة؟ هل سيكون هناك تعاون أم منافسة أم حرب؟



الشكل 10.6: إذا تطورت الحياة بشكل مستقل عند نقاط متعددة في الزمكان (الأماكن والأزمان) وبدأت في استعمار الفضاء، فسيحتوي الفضاء على شبكة من البيوسفيرات الكوزموسية المتوسعة، يشبه كل منها الجزء العلوي من الكأس الزجاجية في الشكل 7.6. ويمثل الجزء السفلي من كل بيوسفير المكان والزمن اللذين بدأ فيهما الاستعمار. وتتوافق الكأس الزجاجية المعتمة وشبه الشفافة مع الاستعمار بنسبة 50% و نسبة 100% من سرعة الضوء، على التوالي، والتقاطع يشير إلى المناطق التي تلتقي فيها الحضارات المستقلة.

تمكن الأوروبيون من التغلب على إفريقيا والأمريكتين لأنهم امتلكوا تكنولوجيا فائقة. على النقيض من ذلك، من المعقول أنه قبل وقت طويل من التقاء حضارتين من الذكاء الفائق، ستصل تكنولوجياتهما إلى المستوى نفسه، مقيدة فقط بحدود قوانين الفيزياء. هذا يجعل من غير المرجح أن يتمكن أحد الذكاءين الخارقين من التغلب على الآخر بسهولة حتى لو أراد ذلك. إضافة إلى ذلك، إذا تطورت أهدافها لتكون متجانسة نسبياً، فربما لا يكون لديها سبب للرغبة في الغزو أو الحرب. مثلاً، إذا كانتا تحاولان إثبات أكبر عدد ممكن من النظريات الجميلة وابتكار خوارزميات ذكية قدر الإمكان، فيمكنهما ببساطة تقاسم اكتشافاتهما وتصبح كلتا الحضارتين أفضل حالاً. ففي المحصلة، تختلف المعلومات اختلافاً كبيراً عن الموارد التي عادة ما يتقاتل البشر عليها، إذ يمكنك في الوقت نفسه منحها عنها والاحتفاظ بها.

وقد يكون لبعض الحضارات الآخذة في التوسع أهدافا غير قابلة للتغيير في الأساس، مثل تلك الموجودة في عقيدة طائفية أو فيروس منتشر. ولكن، من المعقول أيضاً أن تكون بعض الحضارات المتقدمة أشبه بالبشر المتفتحين - المستعدين لتعديل أهدافهم عند مجابهتها بحجج مقنعة. إذا التقت حضارتان منهما، سيكون هناك صدام لكن ليس بين الأسلحة بل بين الأفكار، حيث يسود الأكثر إقناعا وتنتشر أهدافه بسرعة الضوء عبر المنطقة التي تسيطر عليها الحضارة الأخرى. إن احتواء Assimilating جيرانك هو استراتيجية توسع أسرع من الاستيطان، نظراً لأن مجال تأثيرك قد ينتشر بالسرعة التي تتحرك بها الأفكار (سرعة الضوء باستخدام الاتصالات)، في حين أن الاستيطان الفعلي يتقدم أبطأ من سرعة الضوء. لن يتم فرض هذا الاحتواء مثل ذلك الاحتواء سيء الذكر الذي وظّفه البورغ في مسلسل ستار تريك، ولكنه تطوعي يعتمد على التفوق المقنع للأفكار؛ مما يجعل الحضارة المُحتواة في وضع أفضل.

رأينا أن الكون في المستقبل قد يحتوي على نوعين من الفقاعات سريعة التوسع: حضارات مُتوسّعة وفقاعات الموت التي تتسع بسرعة الضوء وتجعل الفضاء غير صالح للسكنى عن طريق تدمير جميع جسيماتنا الأولية. وهكذا يمكن أن تواجه حضارة طموحة ثلاثة أنواع من المناطق: مناطق غير مأهولة، وفقاعات الحياة، وفقاعات الموت. وإذا كانت تخشى حضارات متنافسة غير متعاونة، فإن لديها حافزاً قوياً لإطلاق خطة «استيلاء على الأراضي» سريعة والاستقرار في المناطق غير المأهولة قبل قيام المنافسين بذلك. غير أن لديها الحافز التوسعي نفسه حتى لو لم تكن هناك حضارات أخرى، وذلك ببساطة لضرورة اكتساب الموارد قبل أن تبتعد بها الطاقة المعتمة وتصبح غير قابلة للوصول إليها. لقد رأينا للتو كيف أن الالتقاء المفاجئ بحضارة آخذة في التوسع قد يكون أفضل أو أسوأ من الالتقاء المفاجئ بحيز غير مأهول، وهذا يتوقف على مدى تعاون وانفتاح هذا الجار. ولكن، من الأفضل أن تلتقي بأي حضارة توسعية (حتى تلك التي تحاول أن تحول حضارتك إلى مشابك ورقية) من أن تلتقي بفقاعة موت، ستستمر في التوسع بسرعة الضوء بغض النظر عما إذا كنت تحاول مواجهتها أم التفاهم معها. حمايتنا الوحيدة ضد فقاعات الموت هي الطاقة المعتمة، التي تمنع تلك البعيدة من الوصول إلينا، لذلك إذا كانت فقاعات الموت شائعة بالفعل، فإن الطاقة المعتمة ليست عدونا بل صديقنا.

هل نحن وحيدون؟

كثير من الناس يعتبرون أن هناك حياة متقدمة في معظم أنحاء كوننا. لذا فإن الانقراض البشري لا يهم كثيراً من منظور كوزموسي. ففي المحصلة، لماذا يجب أن نشعر بالقلق إزاء القضاء على أنفسنا إذا كانت بعض الحضارات الملهمة التي تشبه حضارة المسلسل ستار ترك ستدخل قريباً نظامنا الشمسي وتعيد بثّ الحياة فيه، وربما حتى باستخدام تكنولوجيتها المتقدمة لإنعاشنا وإعادتنا إلى الحياة؟ وأعتقد أن فرضية ستار تريك هذه خطيره، لأنها قد تطمئننا بشعور زائف بالأمان مما يجعل حضارتنا غير مبالية ومتهورة. في الواقع، أعتقد أن هذا الافتراض بأننا لسنا وحدنا في كوننا ليس خطيراً فحسب، بل قد يكون خطأ أيضاً.

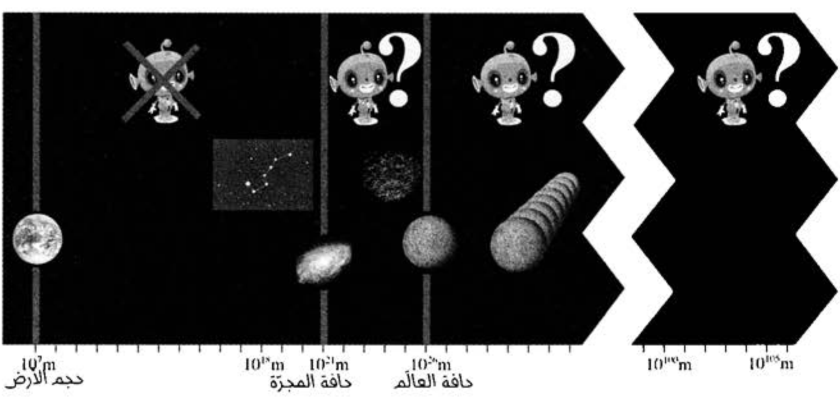
هذه وجهة نظر الأقلية،* وأنا نفسي قد أكون مخطئاً. لكن على الأقل احتمال لا يمكننا حالياً نقضه، الأمر الذي يلقي علينا بحتمية أخلاقية أن نعمل بحرص وألا ندفع بحضارتنا إلى الانقراض.

عندما أُلقي محاضرات حول الكوزمولوجيا (علم الكونيات)، أطلب إلى الجمهور في كثير من الأحيان أن يرفعوا أيديهم إذا كانوا يعتقدون أن هناك حياة ذكية في مكان آخر من كوننا (منطقة الفضاء التي وصلنا منها الضوء حتى الآن خلال 13.8 بليون سنة منذ الانفجار الكبير). وبكل تأكيد، يرفع الجميع تقريباً أيديهم، من رياض الأطفال إلى طلبة الجامعات. وعندما أسأل لماذا؟ فإن الإجابة الأساسية التي أحصل عليها هي أن كوننا ضخم جدّاً بحيث يجب أن تكون هناك حياة في مكان ما، على الأقل من الناحية الإحصائية، دعنا نلقِ نظرة فاحصة على هذه الحجة ونبين ضعفها.

يتعلق الأمر برقم واحد: المسافة النموذجية بين الحضارة في الشكل 10.6 وأقرب جار لها. إذا كانت هذه المسافة أكبر بكثير من 20 بليون سنة ضوئية، يجب أن نتوقع أن نكون وحدنا في كوننا (هذا الجزء من الفضاء الذي وصلنا منه الضوء خلال 13.8 بليون سنة منذ الانفجار الكبير)، وأننا لن نتواصل أبدا مع كائنات فضائية. إذاً، فماذا نتوقع لهذه المسافة؟ نحن جاهلون تماماً، هذا يعني أن المسافة إلى جارتنا هي في حدود 1000...000 متر، حيث يمكن أن يكون العدد الإجمالي للأصفار بشكل معقول 21 ,22 ,23.... أو 100 أو 101 أو 102 أو أكثر - ولكن ربما لا يكون أصغر من 21. نظراً لأننا لم نر بعد أدلة دامغة على الكائنات الفضائية (انظر الشكل 11.6). وكي تكون أقرب حضارة مجاورة لنا داخل كوننا، الذي يبلغ قطر نصف قطره نحو 10^{26} متر، لا يمكن أن يتجاوز عدد الأصفار 26 صفرا، واحتمال أن يقع عدد الأصفار في نطاق ضيق يتراوح بين 22 و 26 هو احتمال صغير إلى حد ما. ولهذا السبب أعتقد أننا وحدنا في كوننا.

* لكن، توصل جون غريبين John Gribbin إلى استنتاج مماثل في كتابه المنشور في عام 2011 **وحيدون في الكون** ***Alone in the Universe*** ولِطيفٍ من وجهات النظر المثيرة للاهتمام حول هذا السؤال، أوصي أيضاً بكتاب بول ديفيز المنشور في عام 2011 بعنوان **الصمت الرهيب** ***Eerie Silence***.

هذا وأقدم تبريراً مفصلاً لهذه الحجة في كتابي **كوننا الرياضياتي** ***Our Mathematical Universe***، لذلك لن أعيد تكرار ذلك هنا. لكن السبب الأساسي وراء جهلنا بمسافة الجار هذه هو أننا بدورنا جاهلون باحتمالات نشوء حياة ذكية في مكان معين، كما أشار عالم الفلك الأمريكي فرانك دريك Frank Drake، يمكن حوسبة هذا الاحتمال بضرب احتمال وجود بيئة صالحة للسكنى (مثل كوكب مناسب)، في احتمال أن تتشكل الحياة هناك، في احتمال تطور هذه الحياة لتصبح ذكية. عندما كنت طالب دراسات عليا، لم أكن أعرف قيمة هذه الاحتمالات الثلاثة. ولكن بعد العقدين الماضيين من الاكتشافات المثيرة للكواكب التي تدور حول النجوم الأخرى، يبدو الآن من المرجح أن الكواكب الصالحة للحياة وفيرة بالبلايين في مجرتنا وحدها. ولكن، احتمال تطور الحياة وبعد ذلك ذكاء لا يزال غير مؤكد جدا: بعض الخبراء يعتقد أن أحدهما أو كليهما أمر لا مفر منه ويحدث في معظم الكواكب المعتادة، في حين يعتقد البعض الآخر أن أحدهما أو كليهما نادر جدّاً بسبب واحد أو أكثر من أحداث عنق الزجاجة التطورية Evolutionary bottlenecks التي تتطلب ضربة حظ لاجتيازها. تتضمن بعض أحداث عنق الزجاجة المقترحة مشكلات الدجاج والبيضة في المراحل المبكرة من حياة تتكاثر ذاتيا: مثلاً، لبناء ريبوسوم Ribosome -الآلة الجزيئية المعقدة جدّاً التي تقرأ الشيفرة الوراثية وتُخلّق بروتيناتنا- تحتاج خلية حديثة إلى ريبوسوم آخر، وليس من الواضح ما إذا كان بإمكان الريبوسوم الأول أن يتطور تدريجيّاً من شيء أبسط.[10] وتتضمن أحداث عنق الزجاجة المقترحة تطوّر ذكاء أعلى، مثلاً، على الرغم من أن الديناصورات حكمت الأرض لأكثر من 100 مليون عام، أي أكثر من ألف ضعف زمن وجود الإنسان الحديث، فلا يبدو أن التطور يدفع حتماً نحو ذكاء أعلى واختراع التلسكوبات أو الحواسيب.



الشكل 11.6: هل نحن وحدنا؟ يقترح عدم اليقين الهائل حول كيفية تطور الحياة والذكاء أن حضارة جارنا الأقرب في الفضاء من المعقول أن تقع في أي مكان على طول المحور الأفقي أعلاه، أي من غير المحتمل أن تكون في النطاق الضيق بين حافة مجرتنا (نحو 10^{21} متر) حافة كوننا (نحو 10^{26} متر). إذا كان الجار أقرب بكثير من هذا النطاق، فلابد من وجود العديد من الحضارات المتقدمة الأخرى في مجرتنا التي ربما كنا سنراها؛ مما يقترح أننا في الواقع وحدنا في عالمنا.

البعض يعارضون حجتي بقولهم: نعم، قد تكون الحياة الذكية نادرة جدّاً. لكنها في الحقيقة ليست كذلك - فمجرتنا تعج بالحياة الذكية التي لا يلاحظها علماء التيار الرئيسي. ربما زارت الكائنات الفضائية الأرض بالفعل، كما يزعم عشاق الأطباق الطائرة. وربما لم تزر الكائنات الفضائية الأرض. لكنهم موجودون في الخارج وهم يختبئون عنا عمداً (وقد أطلق عالِم الفلك الأمريكي جون أيه. بول John A. Ball على هذا "فرضية حديقة الحيوان" Zoo hypothesis، وتناولها كتاب الخيال العلمي الكلاسيكي **صانع النجم** ***Star Maker*** للمؤلف أولاف ستابليدون Olaf Stapledon)، أو ربما يكونون موجودين هناك دون أن يختبئوا عن عمد: إنهم ببساطة غير مهتمين باستيطان الفضاء أو مشاريع هندسية كبيرة يمكننا رصدها.

بالتأكيد، نحن بحاجة إلى أن نظل مستعدين لجميع هذه الاحتمالات، ولكن نظراً لعدم وجود دليل مقبول عموماً على أي منها، نحتاج أيضاً إلى أن نأخذ البديل على محمل الجد: أننا وحدنا. إضافة إلى ذلك، أعتقد أنه لا ينبغي لنا أن نقلل من تنوع الحضارات المتقدمة من خلال افتراض أنها تشترك جميعاً في الأهداف التي تجعلها تظل مخفية: لقد رأينا آنفا أن الحصول على الموارد هدف طبيعي تماماً **لأي** حضارة، وأننا سنلاحظ مثل ذلك، كل ما يتطلبه الأمر هو أن تقرر حضارة واحدة استيطان كل ما في وسعها بشكل علني ومن ثم تبتلع مجرتنا وما وراءها. وفي مواجهة حقيقة وجود الملايين من الكواكب الشبيهة بالأرض والقابلة للسكنى في مجرتنا والتي يزيد عمرها عن عمر الأرض بلايين السنين؛ مما يتيح وقتاً كبيراً للسكان الطموحين لاستيطان المجرة، لذلك لا يمكننا استبعاد التفسير الأكثر وضوحاً: أن أصل الحياة يتطلب ضربة حظ وهكذا لا تكون جميعها مأهولة.

إذا **لم** تكن الحياة نادرة بعد، فقد نعرف ذلك قريباً. تبحث المسوحات الفلكية الطموحة في أجواء الكواكب الشبيهة بالأرض بحثاً عن دليل على الأكسجين الذي تنتجه الحياة. بالتوازي مع هذا البحث عن **أي** حياة، تم تعزيز البحث عن الحياة **الذكية** في الآونة الأخيرة بمشروع المُحْسِن الروسي يوري ميلنر Yuri Milner الذي تبلغ تكلفته 100 مليون دولار: سماع الإنجاز الخارق Breakthrough Listen.

من المهم ألا نركّز على السمات الإنسانية عند البحث عن حياة متقدمة: إذا اكتشفنا حضارة خارج كوكب الأرض، فربما صارت ذكية بالفعل. فكما وصفها مارتن ريس Martin Rees في مقال حديث: «يُقاس تاريخ الحضارة التكنولوجية الإنسانية بالقرون - وربما لا يحتاج الأمر إلا إلى قرن واحد أو اثنين قبل أن يتم التغلب البشر أو تجاوزهم من قبل ذكاء غير عضوي، الذي سيستمر بعد ذلك، ويواصل التطور، لبلايين السنين... سيكون من غير المرجح أن «نرصده» في فترة زمنية قصيرة عندما يتخذ شكلا عضويا».[11] وأنا أتفق مع استنتاج جاي أولسون في ورقة الاستيطان الفضائية المذكورة آنفا: «إننا ننظر إلى احتمال استخدام الذكاء المتقدم لموارد الكون لمجرد ملء الكواكب الأرضية الشبيهة ببساطة بالإصدارات المتقدمة من البشر كنقطة نهاية غير محتملة

لتطوّر التكنولوجيا». لذا عندما تتخيل الكائنات الفضائية، لا تفكر في كائنات خضراء صغيرة ذات ذراعين وساقين، لكن تخيل حياة فائقة الذكاء ترتحل في الفضاء كتلك التي استكشفناها سابقاً في هذا الفصل.

على الرغم من أنني مؤيد قوي لجميع عمليات البحث المستمرة عن حياة خارج كوكب الأرض، والتي تسلط الضوء على واحد من أكثر الأسئلة الرائعة في العلوم، فإنني آمل سرا أن تفشل جميعها ولا تجد شيئا! إن عدم التوافق الواضح بين وفرة الكواكب الصالحة للسكنى في مجرتنا وعدم وجود زوار فضائيين، المعروف أيضاً **بمفارقة فيرمي *Fermi paradox***، يشير إلى وجود ما يسميه الاقتصادي روبن هانسون "المُرشِّح الكبير" Great Filter، وهو حاجز تطوري/تكنولوجي في مكان ما على طريق النمو من المادة غير الحية إلى حياة تستوطن الفضاء. إذا اكتشفنا تطوراً مستمراً في مكان آخر، فهذا سيقترح بأن الحياة البدائية ليست نادرة، وأن الحاجز يكمن بعد المرحلة الحالية من تطور الإنسان - ربما لأن استيطان الفضاء أمر مستحيل أو لأن جميع الحضارات المتقدمة تقريباً تُدمِّر نفسها بنفسها قبل أن تتمكن من السفر في الكوزموس. لذا فإنني آمل ألا تجد جميع عمليات البحث عن حياة فضائية خارج كوكب الأرض شيئا: وهذا يتسق مع السيناريو حيث يكون فيه تطور حياة ذكية أمرا نادرا ولكن نكون نحن البشر محظوظين، بحيث يكون حاجز الطريق وراءنا وتكون لنا إمكانات مستقبلية غير عادية.

توقعات مستقبلية

حتى الآن استكشفنا في هذا الكتاب تاريخ الحياة في عالمنا، منذ بداياتها المتواضعة منذ بلايين السنين إلى العقود الآجلة الباهرة المحتملة في بلايين السنين من الآن. إذا تسبب تطويرنا الحالي للذكاء الاصطناعي في نهاية المطاف في انفجار ذكاء واستيطانات فضاء مُثلى، فسيكون ذلك انفجاراً بالمعنى الكوزموسي الحقيقي: فبعد إنفاق بلايين السنين ككوزموس لا مبالٍ خالٍ من الحياة، تنفجر الحياة فجأة على الساحة الكوزموسية كموجة انفجار كروية تتوسع بسرعة قريبة من سرعة الضوء، ولا تتباطأ أبداً، وتُشعل كل شيء في طريقها مطلقة شرارة الحياة.

هذه الآراء المتفائلة حول أهمية الحياة في مستقبلنا الكوزموسي عبّر عنها ببلاغة العديد من المفكرين الذين التقيناهم في هذا الكتاب. نظراً لأن كتّاب الخيال العلمي يتم تجاهلهم في كثير من الأحيان بوصفهم حالمين رومانسيين غير واقعيين، فقد كان من المفارقات أن معظم كتاب الخيال العلمي والكتابة العلمية متشائمون جدّاً بخصوص استيطان الفضاء في ضوء الذكاء الفائق. مثلاً، رأينا كيف يصبح السفر بين المجرات أسهل بكثير بمجرد انتقال الأفراد والكيانات الذكية الأخرى في شكل رقمي؛ مما يجعلنا ربما نتحكم في مصيرنا ليس فقط في نظامنا الشمسي أو في مجرة درب التبانة، ولكن أيضاً في الكوزموس.

فيما سبق أخذنا بعين الاعتبار احتمال أننا الحضارة التكنولوجيا الفائقة الوحيدة في كوننا. دعونا نقضِ بقية هذا الفصل في استكشاف هذا السيناريو، والمسؤولية الأخلاقية الضخمة التي ينطوي عليها. هذا يعني أنه بعد 13.8 بليون سنة، وصلت الحياة في كوننا إلى مفترق الطرق، لتواجه خيارا بالازدهار في جميع أنحاء الكون أو الانقراض. إذا لم نستمر في تحسين تكنولوجياتنا، فلن يكون السؤال هو ما إذا كانت البشرية *ستنقرض*، ولكن *كيف*. ما الذي سيقضي علينا أولا - كويكب، أم بركان فائق، أم حرارة حارقة من الشمس المتقدمة في العمر، أم كوارث أخرى (انظر الشكل 1.5)؟ بمجرد رحيلنا، ستستمر الدراما الكونية التي تنبأ بها فريمان دايسون لكن من دون متفرجين: انهيار الكون، احتراق النجوم، تلاشي المجرات، تلاشي الثقوب السوداء، كل منها ينهي حياته بانفجار ضخم ينطلق أكثر من مليون ضعف طاقة القنبلة القيصر Tsar Bomba، أقوى قنبلة هيدروجينية بنيت على الإطلاق. وعلى حد تعبير فريمان: «سوف يضيء الكون المتسع البارد بألعاب نارية عرضية لفترة طويلة جدّاً». للأسف، ستهدر متعة الألعاب النارية هذه مع عدم وجود أحد للاستمتاع بها.

ومن دون تكنولوجيا، فإن انقراضنا البشري بات وشيكاً في السياق الكوني من عشرات البلايين من السنين؛ مما يجعل الدراما الكاملة للحياة في عالمنا مجرد لمحة قصيرة وعابرة من الجمال والمرح والمغزى في أبدية لا معنى لها تقريباً لا يشهدها أي شخص. يالها من فرصة ضائعة! وإذاً، بدلاً من تجنب التكنولوجيا، اخترنا احتضانها، وزدنا الرهان: فإننا سنكتسب القدرة على كل من الحياة والبقاء على قيد الحياة، وعلى حد سواء الانقراض في وقت أسرع بفعل التدمير الذاتي بسبب سوء التخطيط (انظر: الشكل 1.5). وأنا أُصوّت لاحتضان التكنولوجيا، والمضي قدماً مؤمنين بما نبنيه، ولكن بحذر وبصيرة وتخطيط دقيق.

بعد 13.8 بليون سنة من التاريخ الكوني، وجدنا أنفسنا في عالم جميل بشكل مذهل، انتشرت فيه الحياه من خلالنا نحن البشر وصار يعي نفسه. لقد رأينا أن إمكانات الحياة المستقبلية في عالمنا أعظم من أغرب أحلام أسلافنا، وإن كانت محفوفة باحتمال حقيقي بالقدر نفسه من أن الحياة الذكية قد تنقرض بشكل دائم. هل ستحقق الحياة في عالمنا إمكاناتها أو تبددها؟ هذا يعتمد إلى حد كبير على ما نفعله نحن البشر الموجودين على قيد الحياة اليوم خلال حياتنا. وأنا متفائل بأننا نستطيع أن نجعل مستقبل الحياة رائعاً حقّاً إذا اتخذنا الخيارات الصحيحة، ماذا يجب أن نريد، وكيف يمكننا تحقيق هذه الأهداف؟ دعنا نقضِ بقية الكتاب في استكشاف بعض من أكثر التحديات صعوبة وما يمكننا القيام به حيالها.

الخلاصة:

- مقارنة بالفترات الزمنية الكونية التي تقدر ببلايين السنين، فإن انفجار الذكاء يعد حدثاً مفاجئاً إذ تتسارع فيه التكنولوجيا في الوصول إلى قمة تكنولوجية لا يحدّها سوى قوانين الفيزياء.
- هذه القمة التكنولوجية أعلى بكثير من التكنولوجيا الحالية؛ مما يسمح لكمية معينة من المادة بتوليد نحو عشرة بلايين ضعف الطاقة (باستخدام السفاليرون أو الثقوب السوداء)، وتخزين 12-18 ضعف المعلومات، أو إجراء حوسبات أسرع 31–41 ضعف- أو ليتم تحويلها إلى أي شكل آخر مرغوب فيه من المادة.
- لن تستخدم الحياة الخارقة الذكاء مواردها الحالية بمثل الاستخدام الفعال جدّاً فحسب، بل ستتمكن أيضاً من تنمية البيوسفير الحالي بنحو 32 ضعف الحجم، من خلال اكتساب المزيد من الموارد عبر الاستيطان الكوني بسرعة تقارب سرعة الضوء.
- الطاقة المعتمة تحد من التمدد الكوني للحياة الخارقة كما أنها تحميها من فقاعات الموت البعيدة أو الحضارات المعادية. وتهديد الطاقة المعتمة بتمزيق الحضارات الكونية، يحفز مشاريع الهندسة الكونية الضخمة، بما في ذلك بناء ثقب دودي إذا كان هذا ممكناً.
- من المرجح أن تكون السلعة الرئيسية المشتركة أو المتداولة عبر المسافات الكونية هي المعلومات.
- فيما عدا الثقوب الدودية، تلقي حدود سرعة الضوء بتحديات عويصة على التنسيق والتحكم عبر الحضارة الكونية. وقد يُحفِّز مركزٌ رئيسي بعيد «العُقد» فائقة الذكاء على التعاون إما من خلال المكافآت أو بالتهديدات، مثلا من خلال نشر حراس ذكاء اصطناعي محليين مبرمجين لتدمير العقد بتفجير سوبرنوفا أو كوازار إذا لم يُمتثَل للقوانين.
- قد ينتج عن تصادم حضارتين متوسعتين امتثال واحدة لثقافة الأخرى Assimilation أو تعاون أو حرب، وإن كانت الحرب أقل احتمالاً مما هي في حضارات اليوم.
- على الرغم من الاعتقاد الشائع بالعكس، فمن المعقول تماماً أننا شكل الحياة الوحيد القادر على جعل عالمنا المرصود ينبض بالحياة في المستقبل.
- إذا لم نُحسّن تقنيتنا، فالمسألة ليست ما إذا كانت البشرية ستنقرض، ولكن كيف: هل سيقضي علينا كويكب، أم بركان فائق، أم حرارة حارقة للشمس المتقدمة في العمر، أم كوارث أخرى؟
- إذا حافظنا على تحسين تكنولوجياتنا بعناية وبعد نظر وتخطيط كافٍ لتجنب المزالق، فإن للحياة القدرة على الازدهار على الأرض وخارجها لعدة بلايين من السنين، بما هو أبعد من أغرب أحلام أسلافنا.

الفصل 7

الأهداف

لا يكمن سر الوجود الإنساني في البقاء على قيد الحياة فقط، ولكن في تحديد شيء للعيش من أجله.

فيودور دوستويفسكي، الأخوان كارامازوف

الحياة عبارة عن رحلة لا وجهة.

رالف والدو ايمرسون

إذا كان عليّ أن أخص في كلمة واحدة ما هي أكثر التناقضات الذكاء الاصطناعي استعصاء، فستكون "الأهداف" Goals: هل يجب أن نعطي الذكاء الاصطناعي أهدافا، وإذا كان الأمر كذلك، فأهداف من؟ وكيف يمكن أن نعطي الذكاءَ الاصطناعي أهدافا؟ وهل يمكننا التأكد من الحفاظ على هذه الأهداف حتى لو صار الذكاء الاصطناعي أكثر ذكاءً؟ هل يمكن أن نغير أهداف ذكاء اصطناعي أذكى منا؟ ما هي أهدافنا النهائية؟ هذه الأسئلة ليست صعبة فحسب، بل هي مهمة أيضاً لمستقبل الحياة: إذا كنا لا نعرف ما نريد، يقل احتمال حصولنا عليه، وإذا تنازلنا عن السيطرة لآلات لا تشاركنا أهدافنا، إذاً ربما نحصل على ما لا نريده.

فيزياء: أصل الأهداف

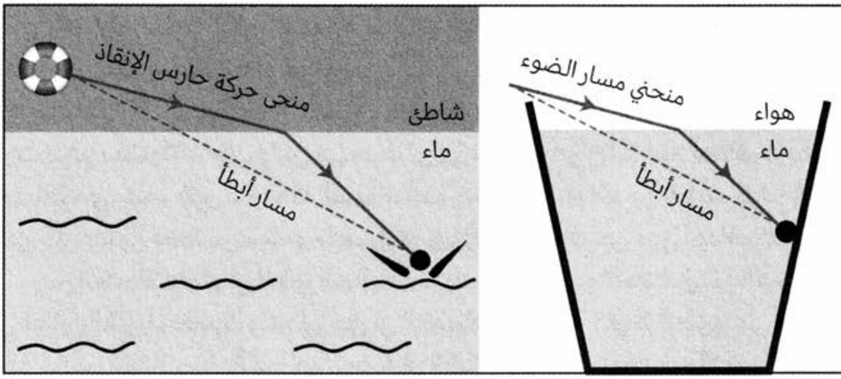
لتسليط الضوء على هذه الأسئلة، دعونا نستكشف الأصل الأقصى للأهداف. عندما ننظر من حولنا إلى العالَم، فإن بعض العمليات تبدو كأنها **موجّهة نحو هدف** ***Goal-oriented*** في حين لايبدو البعض الآخر كذلك. خذ بعين الاعتبار مثلاً، عملية كرة قدم تُركل من أجل

الفوز بالمباراة. سلوك الكرة نفسها لا يبدو موجّها نحو هدف، والطريقة الأكثر اقتصادية علميا هي تفسيره باستخدام قوانين نيوتن للحركة، كرد فعل على ركلة. من ناحية أخرى، الطريقة الأكثر اقتصادية علميا لتفسير سلوك اللاعب ليست الطريقة الميكانيكية من حيث ذرات تدفع بعضها البعض، ولكن من حيث امتلاكه لهدف هو زيادة عدد أهداف فريقه إلى الحد الأقصى. كيف ظهر مثل هذا السلوك الموجّه نحو الأهداف من فيزياء كوننا المبكر، الذي كان يتكون فقط من مجموعة من الجسيمات التي تتقافز من دون أهداف ظاهريا؟

من المثير للاهتمام، أن الجذور النهائية للسلوك الموجّه نحو الأهداف يمكن العثور عليها في قوانين الفيزياء نفسها، وتتجلى حتى في العمليات البسيطة التي لا تنطوي على حياة. إذا أنقذ حارسُ الإنقاذ سباحاً، كما هو موضح في الشكل 1.7. فإننا نتوقع منه ألا يسبح في خط مستقيم، بل أن يجري قليلاً على طول الشاطئ إذ يمكنه أن يكون أسرع منه في الماء، ومن ثم يسبح في خط مائل قليلاً عندما يدخل الماء. نفسر بطبيعة الحال اختياره للمسار على أنه موجّه نحو الهدف، لأن من بين جميع المسارات الممكنة، يختار عمداً المسار الأمثل الذي يجعله يسبح في أسرع وقت ممكن. ولكن شعاعا بسيطا من الضوء ينحني بالمثل عندما يدخل الماء (انظر الشكل 1.7)؛ مما يقلل أيضاً من وقت السفر إلى وجهته! كيف يمكن أن يكون هذا؟

هذا معروف في الفيزياء باسم **مبدأ فيرما** ***Fermat's principle***، الذي وضعه في عام 1662. ويوفر طريقة بديلة للتنبؤ بسلوك أشعة الضوء. ومن المدهش أيضاً أن الفيزيائيين اكتشفوا منذ ذلك الحين أن جميع قوانين الفيزياء الكلاسيكية يمكن إعادة صياغتها بطريقة مماثلة: من بين كل الطرق التي يمكن أن تختار بها الطبيعة القيام بشيء ما، فإنها تفضل الطريقة المُثلى، التي عادةً ما تتلخص في تقليل أو تعظيم كمية ما. هناك طريقتان متعادلتان رياضياتيا لوصف كل قانون فيزيائي: إما مع تسبّب الماضي في حدوث المستقبل، أو مع سعي الطبيعة لأَمْثَلة Optimizing شيء ما. على الرغم من أن الطريقة الثانية عادة لا تُدرّس في مقرارات الفيزياء التمهيدية لأن الرياضيات المتضمنة فيها أكثر صعوبة، غير أنني أشعر أنها أكثر أناقة وعمقاً. إذا كان شخص ما يحاول أَمْثَلة شيء ما (مثلا، النتيجة، أو الثروة أو السعادة)، فسنقوم بطبيعة الحال بوصف سعيه على أنه موجّه نحو هدف. لذا، إذا كانت الطبيعة نفسها تحاول أَمْثَلة شيء ما، فلا عجب إذاً في إمكانية نشوء السلوك الموجّه نحو الهدف: لقد كان مُدمجا بشكل راسخ Hardwired منذ البداية، وفي قوانين الفيزياء ذاتها.

وهناك كمية شهيرة تسعى الطبيعة للوصول بها إلى أقصى قيمة Maximize ألا وهي **الإنتروبية** ***Entropy***، التي تقيس بشكل عام مدى فوضى شيء ما. ينص القانون الثاني للديناميكا الحرارية على أن الإنتروبية تميل إلى الزيادة حتى تصل إلى أقصى قيمة ممكنة لها. تجاهل آثار الجاذبية في الوقت الحالي، تسمى هذه الحالة الفوضوية القصوى الموت الحراري Heat death، الذي يُمثّل توزيع كل شيء في انتظام (تجانس) مثالي ممل، مع عدم وجود تعقيد، لا حياة ولا تغيير. عندما تصب الحليب البارد في القهوة الساخنة، مثلاً يبدو أن مشروبك يسير في اتجاه لا يمكن عكسه نحو هدف موته الحراري، وقبل مضي وقت



الشكل 1.7: لإنقاذ سباح في أسرع وقت ممكن، لن يسبح حارس الإنقاذ في خط مستقيم (الخط المتقطع)، ولكن سيجري أولاً أبعد قليلاً على طول الشاطئ حيث يمكنه أن يتحرك أسرع منه في الماء. وشعاع الضوء ينحني بالمثل عند دخول الماء للوصول إلى وجهته في أسرع وقت ممكن.

طويل، يكون كله مجرد مزيج فاتر متجانس. إذا مات كائن حي، تبدأ الإنتروبية أيضاً في الارتفاع، وقبل مضي وقت طويل، يميل ترتيب جزيئاته إلى أن يصبح أقل تنظيماً.

هدف الطبيعة الظاهري لزيادة الإنتروبية يساعد على تفسير السبب الذي يبدو فيه أن للزمن اتجاهاً مفضلاً؛ مما يجعل الأفلام تبدو غير واقعية إذا تم تشغيلها بالعكس إلى من النهاية إلى البداية: إذا أسقطت كوباً من شراب، فستتوقع أن تتحطم الكأس عند وصولها إلى الأرض وتزيد الفوضى العامّة (الإنتروبية). إذا شاهدت بعد ذلك حالة لاتحطم Unshatter الكأس وعودتها طائرة إلى يديك سليمة (تقليل الإنتروبية)، فمن المحتمل أنك لن تشرب منها، لأنك ستعتقد أنك قد شربت الكفاية بالفعل.

عندما علمت أول مرة عن تقدمنا الثابت نحو الموت الحراري، وجدت الأمر محبطا جدّاً، ولم أكن وحدي في ذلك: فقد كتب اللورد كيلفن Lord Kelvin رائد الديناميكا الحرارية في عام 1841 أن «النتيجة ستكون حتماً حالة من السكون والموت الشاملين»، ومن الصعب أن تجد العزاء في فكرة أن الهدف الطويل الأجل للطبيعة هو تحقيق أقصى قدر من الموت والدمار. لكن أظهرت الاكتشافات الحديثة أن الأمور ليست بهذا السوء. أولا، تتصرف الجاذبية بشكل مختلف عن جميع القوى الأخرى وتسعى جاهدة لجعل عالمنا ليس أكثر تجانسا ومللاً بل أكثر تكتلا وإثارة للاهتمام. لذلك حولت الجاذبية عالمنا المُمل المبكّر، الذي كان متجانساً تماماً، إلى كون اليوم الغامض والمعقد بشكل جميل، الزاخر بالمجرات والنجوم والكواكب. وبفضل الجاذبية، هناك الآن مدى واسع من درجات الحرارة التي تتيح للحياة أن تزدهر من خلال الجمع بين الحار والبارد: نحن نعيش على كوكب دافئ مريح يمتص حرارة شمسية تبلغ 6,000 درجة سيليزية (10,000 درجة فهرنهايت) في حين يبرد عن طريق إشعاع الحرارة الزائدة في الفضاء المتجمد الذي درجة الحرارة ليست سوى 3 درجات سيليزية (5 فهرنهايت) فوق الصفر المطلق.

ثانياً، إن العمل الذي قام به مؤخراً زميلي في المعهد MIT جيريمي إنغلاند Jeremy England وآخرون جلب المزيد من الأخبار الجيدة، فقد بيّنوا أن الديناميكا الحرارية تمنح الطبيعة أيضاً هدف أكثر إلهاماً من الموت الحراري.[1] يُطلق على هذه الهدف مصطلح **التكيّف المدفوع بالتبديد** ***Dissipation-driven adaptation*** بلغة مهوسي العلوم، وهو ما يعني جوهريا أن مجموعات عشوائية من الجسيمات تسعى جاهدة لتنظيم نفسها من أجل استخراج الطاقة من بيئتها بأفضل كفاءة ممكنة (مصطلح "التبديد" Dissipation يعني التسبب في زيادة الإنتروبية، عادة عن طريق التحويل المفيد للطاقة إلى حرارة، وفي كثير من الأحيان عند القيام بعمل مفيد أثناء ذلك). مثلاً، تميل مجموعة من الجزيئات التي تتعرض لأشعة الشمس بمرور الوقت إلى ترتيب نفسها لتصبح أفضل فأفضل في امتصاص أشعة الشمس. بعبارة أخرى، يبدو أن للطبيعة هدفاً مغروساً بداخلها - إنتاج أنظمة ذاتية التنظيم معقدة بشكل متزايد وشبيه بالحياة، وهذا الهدف مغروس في قوانين الفيزياء ذاتها.

كيف يمكننا التوفيق بين هذا الدافع الكوني نحو الحياة والدافع الكوني نحو الموت الحراري؟ يمكن العثور على الإجابة في الكتاب الشهير **ما الحياة؟** ***What's Life*** المنشور في عام 1944 لإروين شرودنغر Erwin Schrödinger، أحد مؤسسي ميكانيكا الكم Quantum mechanics. أشار شرودنغر إلى أن السمة المميزة لنظام حيّ هي أنه يحافظ على أو يقلل من الإنتروبية عن طريق زيادة الإنتروبية المحيطة به. بعبارة أخرى، يحتوي القانون الثاني للديناميكا الحرارية على ثغرة في الحياة: على الرغم من أن مجموع الإنتروبية يجب أن يزداد، فإنه يُسمح بالنقص في بعض الأماكن ما دام يزداد في أماكن أخرى، لذلك تحافظ الحياة على -أو تزيد من تعقيدها- عن طريق جعل بيئتها أكثر فوضى.

بيولوجيا: تطور الأهداف

رأينا للتو الكيفية التي يمكن بها تتبع أصل السلوك الموجّه نحو الهدف وصولا إلى قوانين الفيزياء، التي يبدو أنها تهب الجسيمات هدف ترتيب أنفسها لاستخراج أكبر قدر ممكن من الطاقة من بيئتها. وهناك طريقة رائعة لترتيب الجسيمات لتعزيز هذا الهدف هي عمل نسخ من نفسها، لإنتاج المزيد من ممتصي الطاقة. هناك العديد من الأمثلة المعروفة لمثل هذا التكرار الذاتي الناشئ: مثلاً، يمكن أن تقوم الدوامات المضطربة في سائل بنسخ نفسها، ويمكن لعناقيد من الكرات المجهرية Microspheres دفع الكرات القريبة لتكوين عناقيد متطابقة. في مرحلة ما، صار ترتيب معين من الجسيمات جيداً في نسخ نفسه بحيث صار يمكنه فعل ذلك بشكل لا منتهٍ تقريباً عن طريق استخراج الطاقة والمواد الخام من بيئته. نحن نسمي ترتيب الجسيمات هذا: حياة. ولانزال نجهل الكيفية التي نشأت بها الحياة على الأرض. لكننا نعرف أن أشكال الحياة البدائية كانت موجودة منذ أربعة بلايين سنة.

إذا كان شكل الحياة ينسخ نفسه والنُسخ تفعلُ الشيء نفسه، إذاً سيظل العدد الإجمالي يتضاعف على فترات زمنية منتظمة حتى يتعارض حجم السكان مع قيود الموارد

أو غيرها من المشكلات. وسرعان ما ستنتج المضاعفة المتكررة أعداداً ضخمة: إذا بدأت برقم واحد وضاعفته 300 مرة فقط، فستحصل على كمية تفوق عدد الجسيمات في عالمنا. هذا يعني أنه بعد وقت قصير من ظهور شكل الحياة البدائية الأولى، صارت كميات هائلة من المادةِ حيّةً. في بعض الأحيان لم تكن النُسَخ مثالية، لذا سرعان ما كانت هناك أشكال حياة مختلفة تحاول نسخ نفسها، متنافسة على الموارد المحدودة نفسها. وبدأ التطور الداروين.

إذا كنت تراقب الأرض بهدوء في الوقت الذي بدأت فيه الحياة، فستلاحظ تغييراً هائلاً في السلوك الموجّه نحو الهدف. فبينما بدت الجسيمات في وقت سابق، كما لو كانت تحاول زيادة متوسط الفوضى بطرق مختلفة، يبدو أن أنماط النسخ الذاتي الجديدة المنتشرة في كل مكان كان لها هدف مختلف: ليس التبديد Dissipation ولكن **النسخ *Replication***. أوضح تشارلز داروين بأناقة لماذا: نظراً لأن أكثر آلات النسخ كفاءة تتفوق على الآخرين وتسيطر عليهم، فإن أي شكل عشوائي من الحياة تنظر إليه سرعان ما سيصل إلى الصورة المثلى كي يحقّق هدف النسخ.

كيف يمكن أن يتغير الهدف من التبديد إلى النسخ في حين تظل قوانين الفيزياء كما هي؟ الجواب هو أن الهدف الأساسي (التبديد) لم يتغير، لكنه أدى إلى هدف غائي Instrumental goal، أي هدف فرعي يساعد على تحقيق الهدف الأساسي. خذ الأكل، مثلاً، يبدو أننا جميعا لدينا هدف إشباع شعورنا بالجوع على الرغم من أننا نعرف أن الهدف الأساسي الوحيد للتطور هو النسخ وليس المضغ. ولأن تناول الطعام يساعد على التكاثر، فإن الموت جوعا يعيق إنجاب الأطفال. بالمثل، يساعد النسخ على التبديد، لأن الكوكب الذي يعج بالحياة يكون أكثر فعالية في تبديد الطاقة. لذلك، بمعنى ما، طوّر الكونُ حياتنا لمساعدته على الاقتراب من الموت الحراري بشكل أسرع. إذا سكبت السكر على أرض مطبخك، فيمكنه من حيث المبدأ الاحتفاظ بطاقته الكيميائية المفيدة لسنوات، ولكن إذا ظهر النمل، فسيبددون هذه الطاقة سريعا. وبالمثل، فإن احتياطيات النفط المدفونة في قشرة الأرض كانت ستحتفظ بطاقتها الكيميائية المفيدة لفترة أطول لو أن شكل حياتنا الواقف على قدمين لم يضخها ويحرقها.

اليوم وفيما بين سكان الأرض المتطورين، يبدو أن هذه الأهداف الفعّالة Instrumental اكتسبت مسارها الخاص: على الرغم من أن التطور اختارهم لتحقيق الهدف الوحيد المتمثل في النسخ، فإن الكثير منها يقضي الكثير من وقته في عدم إنتاج ذرية، ولكن على أنشطة مثل النوم والبحث عن الطعام وبناء المنازل والتأكيد على الهيمنة والاقتتال أو مساعدة الآخرين - حتى في بعض الأحيان إلى حد يقلّل من النسخ. وقد فسر علماء النفس التطوري والاقتصاد وعلم الآثار بأناقة لماذا يكون ذلك. بعض علماء الاقتصاد اعتادوا على نمذجة الأفراد كوكلاء منطقيين Rational agents، صناع قرار مثاليين دائما يختارون الفعل الأمثل في السعي لتحقيق هدفهم، ولكن هذا من الواضح أنه غير واقعي. ففي الممارسة العملية، يمتلك هؤلاء الوكلاء ما أطلق عليه الحاصل على جائزة نوبل هيربرت سيمون Herbert Simon "العقلانية

المقيدة" Bounded rationality لأن لديهم موارد محدودة: إن عقلانية قراراتهم محدودة بالمعلومات المتوفرة لديهم، والوقت المتاح للتفكير فيها وأجهزتهم المتاحة للتفكير بها. وهذا يعني أنه عندما يقوم التطور الدارويني بتحسين الكائن الحي للوصول إلى الهدف، فإن أفضل ما يمكن عمله هو تنفيذ خوارزمية تقريبية تعمل جيداً بشكل معقول في السياق المُقيَّد الذي يجد الوكيلُ نفسَه فيه. لقد طبق التطور أمثلة النَسْخ بهذه الطريقة بالتحديد: بدلاً من السؤال في كل موقف عن أي فعل يجب القيام به ليزيد الكائن من عدد ذريته إلى أقصى حد، فإنه ينفذ خليطا من الدوافع: مطبقا قواعد أساسية عادة ما تعمل بشكل جيد. بالنسبة إلى معظم الحيوانات، تشمل هذه الدافع التكاثر والشرب عند العطش، وتناول الطعام عند الجوع وتجنب الأطعمة سيئة الطعم أو الأمور المؤلمة.

أحياناً تفشل القواعد الأساسية هذه فشلا ذريعا في المواقف التي لم يتم تصميم الكائن الحي للتعامل معها، مثلا عندما تأكل الفئران سم الفئران لذيذ المذاق، وعندما تُغرى حشرات العث لدخول مصائد الغراء باستخدام رائحة أنثاها، وعندما تطير الحشرات إلى ضوء الشموع.* ونظراً لأن المجتمع البشري اليوم مختلف تماماً عن القواعد الأساسية التي أمثلتها البيئة التي تطورنا فيها، فلا ينبغي أن نتفاجأ أننا غالباً ما يفشل سلوكنا في إنتاج أقصى حد من الأطفال. ومثلا، يتم تحقيق الهدف الفرعي المتمثل في عدم الموت جوعا جزئياً على شكل رغبة في تناول الأطعمة التي تحتوي على السعرات الحرارية العالية؛ مما يؤدي إلى انتشار وباء السمنة اليوم. وطُبِّق الهدف الفرعي للتكاثر كرغبة في ممارسة الجنس بدلاً من الرغبة في أن يصبح الفرد مانحاً للحيوانات المنوية/البويضات، على الرغم من أنه يمكنه أن ينتج المزيد من الأطفال بأقل جهد.

علم النفس: السعي وراء تحقيق الأهداف والتمرد ضدها

باختصار، إن الكائن الحي هو وكيل العقلانية المُقيَّدة التي لا تسعى إلى تحقيق هدف، ولكن بدلاً من ذلك تتبع القواعد الأساسية لما يجب السعي له وتجنبه. تدرك عقولُنا البشرية أن هذه القواعد الأساسية تطورت على شكل **مشاعر *Feelings***، عادةً (وغالباً دون علمنا بها) تُوجِّه عملية اتخاذ القرار نحو الهدف النهائي للنسخ. فمشاعر الجوع والعطش تحمينا من الجوع والجفاف، ومشاعر الألم تحمينا من إلحاق الضرر بجسمنا، ومشاعر الشهوة تجعلنا نتكاثر، ومشاعر الحب والرحمة تجعلنا الناقلين الآخرين لجيناتنا وجينات أولئك الذين يساعدونها وهلم جرا. موجّهين بهذه المشاعر، يمكن لأدمغتنا أن تقرر بسرعة وبكفاءة ما يجب القيام به دون الاضطرار إلى إخضاع كل خيار لتحليل تفصيلي للتأثيرات القصوى

* من القواعد الأساسية التي تستخدمها العديد من الحشرات التي تطير في خط مستقيم هي افتراض أن الضوء الساطع هو الشمس فتطير بزاوية ثابتة بالنسبة إليها. وإذا كان هذا الضوء لهباً قريباً، فيمكن لهذا مع الأسف أن يخدع الحشرة إلى الطيران نحو دوامة الموت.

له في عدد الذرية التي سينتجها الكائن الحي. للحصول على منظور ذي صلة حول المشاعر وجذورها الفسيولوجية، أوصي بشدة بقراءة كتابات ويليام جيمس William James وأنتونيو داماسيو António Damásio.[2]

من المهم أن نلاحظ أنه عندما تعمل مشاعرنا أحياناً ضد الحصول على طفل، فهذا ليس بالضرورة عن طريق الصدفة أو لأننا نُخدَع: إذ يمكن لعقلنا التمرد على جيناتنا وهدف نسخها المتعمد، مثلا باختيار استخدام موانع الحمل! وتشمل الأمثلة الأكثر تطرفاً على تمرد الدماغ ضد جيناته اختيار الانتحار أو العيش عيشة العزوبية ليصبح كاهناً أو راهباً أو راهبة.

لماذا نختار أحياناً التمرد على جيناتنا وهدف نسخها؟ نحن متمردون بسبب التصميم، بوصفنا وكلاء للعقلانية المُقيَّدة، فنحن مخلصون فقط لمشاعرنا. على الرغم من أن أدمغتنا تطورت لمجرد المساعدة على نسخ جيناتنا، فإن أدمغتنا لم تكن تهتم بهذا الهدف نظراً لعدم وجود أي مشاعر مرتبطة بالجينات. وفي الواقع، خلال معظم تاريخ البشرية، لم يعرف أسلافنا حتى أنهم لديهم جينات. إضافة إلى ذلك، أدمغتنا هي أكثر ذكاءً من جيناتنا بكثير، والآن بعد أن فهمنا هدف جيناتنا (النسخ)، فإننا نجد ذلك أمرا عاديا ويسهل تجاهله. قد يدرك الناس السبب في أن جيناتهم تثير مشاعر الشهوة، ولكن لديهم رغبة قليلة في تربية 15 طفلاً، لذا يختارون اختراق برمجياتهم الوراثية من خلال الجمع بين المكافآت العاطفية للعلاقة الحميمة مع تحديد النسل، قد يدركون لماذا تجعلهم جيناتهم يشتهون الحلوى ولكن لديهم رغبة قليلة في زيادة الوزن، ومن ثم يختارون اختراق برمجياتهم الوراثية البرنامج من خلال الجمع بين المكافآت العاطفية للمشروبات الحلوة صفرية السعرات الحرارية المُحلَّاة بالمحليات الاصطناعية.

مع أن آليات المُكافأة لاختراق النظام قد تحيد عن المسار، كما هي الحال عندما يدمن الناس على الهيروين، فإن مجموعة جيناتنا البشرية نجت حتى الآن على الرغم من أدمغتنا البارعة والمتمردة. ولكن، من المهم أن نتذكر أن السلطة المطلقة هي الآن مشاعرنا، وليس جيناتنا، هذا يعني أن السلوك البشري لم يتم أمثلته بشكل صارم لبقاء جنسنا البشري. في الواقع، نظراً لأن مشاعرنا تطبق القواعد الأساسية غير المناسبة في جميع الحالات، فإن السلوك الإنساني بالمعنى الدقيق للكلمة ليس له هدف محدد واحد على الإطلاق.

الهندسة: تعهيد الأهداف لمصادر خارجية

هل يمكن للآلات أن يكون لديها أهداف؟ لقد أثار هذا السؤال البسيط جدالاً كبيراً، نظراً لأن الأشخاص المختلفين يقصدون أشياء مختلفة، وغالباً ما يرتبط هذا السؤال بموضوعات شائكة، مثل ما إذا كانت الآلات يمكن أن تكون واعية وما إذا كان يمكن أن يكون لديها مشاعر. لكن إذا كنا أكثر عملية وأخذنا السؤال ببساطة ليعني: «هل يمكن للآلات أن تظهر سلوكاً موجهاً نحو الهدف؟»، فإن الجواب واضح: «بالطبع يمكنها ذلك، إذ يمكننا تصميمها بهذه الطريقة!» فنحن نصمم المصيدة بهدف اصطياد الفئران، وغسالات الصحون بهدف تنظيف

الأطباق والساعات بهدف رصد الوقت. عندما تُواجِه آلةً، فإن الحقيقة العملية هي أنها تظهر سلوكا موجّها بالهدف - عادةً ما يكون السلوك الموجه هو ما يهمك: إذا كان صاروخ موجّه بالحرارة يطاردك، فلا يهمك حقّاً ما إذا كان لديه وعي أو مشاعر! إذا كنت لا تزال تشعر بعدم الارتياح من القول إن للصاروخ هدفا حتى لو لم يكن واعيا، يمكنك الآن ببساطة استبدال ذلك بـ "غرض" Purpose عندما أكتب "هدف" Goal - سنتناول الوعي في الفصل التالي.

حتى الآن، فإن معظم ما نُصنّع يعكس فقط تصاميم موجهة نحو الهدف، وليس سلوكا موجّها نحو الهدف: فليس للطريق السريع **سلوك**؛ إنه يقبع هناك فقط. ولكن التفسير الأكثر اقتصادية لوجوده هو أنه تم تصميمه لتحقيق هدف، لذلك حتى هذه التكنولوجيا السالبة Passive تجعل الكون أكثر توجها نحو الأهداف. أما **علم الغائية *Teleology*** فهو تفسير الأشياء من حيث أغراضها بدلاً من أسبابها، لذلك يمكننا تلخيص الجزء الأول من هذا الفصل بالقول إن عالمنا آخذ في التحول إلى المزيد من الغائية.

ليس بإمكان المادة غير الحية أن يكون لها أهداف، على الأقل بهذا المعنى الضعيف، بل يتزايد وجود أهداف لها. إذا كنت تراقب ذرات الأرض منذ تكوين كوكبنا، كنت ستلاحظ ثلاث مراحل من السلوك الموجّه نحو الأهداف:

1. بدا أن كل شيء يركز على التبديد (زيادة الإنتروبية).
2. صار بعض المادة حيّا، وركزت الحياة بدلاً من ذلك على النسخ والأهداف الفرعية لذلك.
3. أُعيد ترتيب جزء سريع النمو من المادة وذلك من قبل الكائنات الحية للمساعدة على تحقيق أهدافها.

كيانات موجهة بالهدف	بلايين الأطنان
5×10^{30} باكتيريا	400
نباتات	400
10^{15} أسماك المياه متوسطة العمق Mesophelagic fish	10
1.3×10^{9} أبقار	0.5
7×10^{9} بشر	0.4
10^{14} نمل	0.3
1.7×10^{6} حيتان	0.0005
كونكريت	100
حديد صلب	20
إسفلت	15
1.2×10^{9} سيارات	2

الجدول 1.7: الكميات التقريبية للمادة على الأرض في كيانات تطوّرت -أو تم تصميمها- بحيث تكون موجَّهة نحو هدف ما. تظهر الكيانات المُصمَّمة هندسيا، مثل المباني والطرق والسيارات، كما لو أنها تتبع مسارا تتغلب فيه أعدادها على الكيانات المتطورة مثل النباتات والحيوانات.

يوضح الجدول 1.7 الكيفية التي نجمت بها البشرية المُهيمنة من منظور الفيزياء: ليس فقط لأننا نحتوي الآن على مادة أكثر من جميع الثدييات الأخرى باستثناء الأبقار (التي تعد عديدة لأنها تخدم أهدافنا المتمثلة في استهلاك لحوم الأبقار ومنتجات الألبان)، ولكن المادة في أجهزتنا وطرقنا ومبانينا ومشاريعنا الهندسية الأخرى يبدو أنها تتبع مساراً ستتغلب فيه على جميع المواد الحية على الأرض. بعبارة أخرى، حتى من دون انفجار ذكاء، فإن معظم المادة الموجودة على وجه الأرض، والتي تظهر خصائص موجهة نحو الأهداف، قد يتم تصميمها قريباً وليس تطويرها.

هذا النوع الثالث الجديد من السلوك الموجّه نحو الأهداف لديه القدرة على أن يكون أكثر تنوعاً مما سبقه: بينما يكون لجميع الكيانات التي تطورت الهدف الأقصى نفسه (النسخ)، فقد يكون للكيانات المُصمَّمة تقريباً أي هدف أقصى، بل حتى أهداف متضادة. تحاول الأفران تسخين الطعام في حين تحاول الثلاجات تبريد الطعام، تحاول المولدات تحويل الحركة إلى كهرباء في حين تحاول المحركات تحويل الكهرباء إلى حركة، تحاول برمجيات الشطرنج القياسية الفوز بالشطرنج. لكن هناك أيضاً برمجيات تتنافس في البطولات بهدف الخسارة في الشطرنج.

هناك اتجاه تاريخي للكيانات المُصمَّمة للحصول على أهداف ليس فقط أكثر تنوعاً، ولكن أيضاً أكثر تعقيداً: فقد صارت أجهزتنا أكثر ذكاءً. صممنا أجهزتنا الأقدم وغيرها من الأعمال الفنية لتحقيق أهداف بسيطة، مثلا المنازل التي تهدف إلى إبقائنا دافئين وغير مبلولين وآمنين. وتعلمنا تدريجيّاً صنع آلات ذات أهداف أكثر تعقيداً، مثل المكانس الكهربائية الآلية والصواريخ ذات الطيار الآلي والسيارات ذاتية القيادة. وقدم لنا التقدم الذي أحرزه الذكاء الاصطناعي مؤخراً أنظمة مثل ديب بلو Deep Blue وواتسون Watson وألفا غو AlphaGo، التي كانت أهدافها الفوز في الشطرنج، والفوز في مسابقات الأسئلة، والفوز في لعبة Go، وهي أنظمة معقدة جدّاً لدرجة يتطلب الأمر معها إتقاناً بشرياً كبيراً ليقدر مدى مهارة هذه الأنظمة.

عندما نصنع آلة لمساعدتنا، قد يكون من الصعب التوفيق بين أهدافها وأهدافنا، مثلاً، قد يلتبس الأمر على مصيدة الفئران فتظن أصابعك قدم فأرا جائعا؛ مما يؤدي إلى نتائج مؤلمة. جميع الآلات هي وكلاء ذات عقلانية مُقيَّدة، وحتى أكثر الأجهزة تطوراً اليوم لديها فهم أقل للعالم من فهمنا، ومن ثم فإن القواعد التي تستخدمها لتوضيح ما يجب القيام به غالباً ما تكون مبسطة جدا. فمصيدة الفئران سعيدة جدّاً بقدح الزناد، إذ ليس لديها أدنى فكرة عن ماهية الفأر، والكثير من الحوادث الصناعية المميتة التي تحدث بسبب الآلات ليس لديها أدنى فكرة عن ماهية الفرد، والحواسيب التي أدت إلى «الانهيار المفاجئ» في وول ستريت في عام 2010 لم يكن لديها أدنى فكرة عن أن ما فعلته ليس بالمنطقي. ومن ثم، يمكن حل العديد من مشكلات مواءمة الأهداف من خلال جعل أجهزتنا أكثر ذكاءً. لكننا نحن أيضاً تعلمنا من بروميثيوس في الفصل الرابع. أن ذكاء آلة متناميا باستمرار يثير تحديات جديدة خطيرة حول كيفية ضمان أن تتشارك الآلات معنا في أهدافنا.

ذكاء اصطناعي ودود: مواءمة الأهداف

كلما زادت الآلات ذكاءً وقوة، صار من المهم أن تتواءم أهدافها مع أهدافنا. ما دمنا لا نبني إلا آلات غبية نسبياً، فإن السؤال لا يتعلق بما إذا كانت الأهداف البشرية ستسود في النهاية، ولكن كم من المشكلات ستتسبب بها هذه الآلات للبشرية قبل أن نكتشف كيفية حل مشكلة مواءمة الأهداف. لكن إذا أُطلق ذكاء فائق في يوم من الأيام، فسيكون الأمر عكس ذلك: لما كان الذكاء هو القدرة على تحقيق الأهداف، فإن الذكاء الاصطناعي الفائق الذكاء -تعريفا- أفضل بكثير في تحقيق أهدافه منا نحن البشر في تحقيق أهدافنا، ومن ثم سيسود الذكاء الاصطناعي الفائق الذكاء. في الفصل الرابع استكشفنا العديد من هذه الأمثلة التي تتضمن بروميثيوس. إذا كنت ترغب في تجربة أهداف جهاز تتفوق على أهدافك الآن، فما عليك سوى تنزيل أحدث برمجيات الشطرنج وحاول التغلب عليها. لن تتمكن أبدا من التغلب على البرمجية، كما تتقادم البرمجية بسرعة...

بعبارة أخرى، **الخطر الحقيقي مع الذكاء *AGI* ليس الخبث بل الكفاءة**. سيكون الذكاء الاصطناعي الفائق الذكاء جيداً جدّاً في تحقيق أهدافه، وإذا لم تتواءم هذه الأهداف مع أهدافنا، فنحن في ورطة. كما أشرت في الفصل الأول، لا يتردد الناس في غمر بيوت النمل عند بناء السدود الكهرومائية، لذلك دعونا لا نضع البشرية في موضع هؤلاء النمل. لذلك، يجادل معظم الباحثين بأنه إذا انتهى بنا المطاف إلى بناء ذكاء فائق، فعلينا التأكد من أنه من النوع الذي أطلق عليه رائد أمن الذكاء الاصطناعي اليعازر يودكوفسكي Eliezer Yudkowsky اسم "الذكاء الاصطناعي الودود" Friendly AI: الذكاء الاصطناعي الذي تتواءم أهدافه مع أهدافنا.[3]

إن حل كيفية مواءمة أهداف الذكاء الاصطناعي الفائق مع أهدافنا ليس بالأمر المهم فحسب، ولكنه صعب أيضاً. في الواقع، إنه حالياً مشكلة لم تحل بعد. وتنقسم إلى ثلاث مشكلات فرعية صعبة، كل منها هي موضوع بحث يعمل عليه علماء الحاسوب والمفكرون الآخرون:

1. جعل الذكاء الاصطناعي **يتعلم** أهدافنا
2. جعل الذكاء الاصطناعي **يتبنّى** أهدافنا
3. جعل الذكاء الاصطناعي **يتمسك** بأهدافنا

دعنا نستكشفها واحدا بعد الآخر، ونؤجل مسألة ماذا نعني بـ "أهدافنا" إلى القسم التالي. لتعلُّم أهدافنا، يجب على الذكاء الاصطناعي أن يفهم ليس ما نفعله، ولكن لماذا نفعل ذلك. نحن البشر نتمكن من ذلك دون عناء لدرجة يسهل معها أن ننسى مدى صعوبة المهمة بالنسبة إلى حاسوب، ومدى سهولة إساءة فهمها. إذا طلبت إلى سيارة مستقبلية ذاتية القيادة أن تأخذك إلى المطار بأسرع وقت ممكن ونفذت السيارة كلامك بالمعنى الحرفي، فستصل إلى هناك مطاردا بالمروحيات ومغطى بالقيء. وإذا صرخت معترضا: «هذا ليس ما أردت!»، فبإمكانه الإجابة إجابةً مبرّرة: «هذا ما طلبته أنت». يتكرر الأمر نفسه في العديد من القصص الشهيرة. في الأسطورة اليونانية القديمة، طلب الملك ميداس أن

يتحول كل ما يلمسه إلى ذهب، ولكنه شعر بخيبة أمل عندما منعه هذا من تناول الطعام، وازدادت خيبته أكثر عندما حول ابنته عن غير قصد إلى ذهب. وفي القصص التي يمنح الجني فيها ثلاث رغبات، هناك العديد من الصيغ للرغبتين الأوليين. لكن الرغبة الثالثة هي نفسها دائماً تقريباً: «الرجاء التراجع عن أول رغبتين، لأن هذا ليس ما أردته حقاً».

جميع هذه الأمثلة توضّح أنه من أجل التعرف على ما يريده الناس حقاً، لا يمكنك مجرد أخذ ما يقولون بالمعنى الحرفي. أنت بحاجة أيضاً إلى نموذج تفصيلي للعالم، بما في ذلك التفضيلات المشتركة العديدة التي نميل إلى عدم الإفصاح عنها لأننا نعتبرها واضحة، مثل أننا لا نحب القيء أو تناول الذهب. بمجرد أن يكون لدينا مثل هذا النموذج العام، يمكننا في كثير من الأحيان معرفة ما يريده الناس حتى لو لم يخبرونا به، وذلك ببساطة عن طريق مراقبة سلوكهم الموجّه نحو الأهداف. في الواقع، فإن أطفال المنافقين عادة يتعلمون من مشاهدة أفعال آبائهم أكثر مما يتعلمونه من سماع كلامهم.

يحاول باحثو الذكاء الاصطناعي حالياً جاهدين تمكين الآلات من استنتاج أهدافٍ من السلوك، وسيكون ذلك مفيداً أيضاً قبل وقت طويل من ظهور أي ذكاء فائق في المشهد. مثلاً، قد يُثمِّن رجل متقاعد أن يتمكن روبوته الخاص بكبار السن من فهم ما يهتم به الرجل ببساطة من خلال ملاحظة تصرفاته، حتى يُجنَّب إزعاج شرح كل شيء بالكلمات أو برمجة الحاسوب. وأحد التحديات هو إيجاد طريقة جيدة لترميز الأنظمة الاعتباطية للأهداف والمبادئ الأخلاقية في حاسوب، وهناك تحد آخر هو صنع آلات يمكنها تحديد أي نظام معين يتناسب بشكل أفضل مع السلوك الذي يرصده النظام.

والمقاربةُ المحبّبة حالياً لمواجهة التحدي الثاني تُعرف في لغة المختصين المهووسين Geek-talk **بالتعلم بالتعزيز العكسي *Inverse reinforcement learning***، وهو محور التركيز الرئيسي لمركز أبحاث جديد في بيركلي أطلقه ستيوارت راسل Stuart Russell. لنفترض، مثلاً، أن ذكاء اصطناعيا يراقب رجل إطفاء يهرع إلى داخل مبنى محترق وينقذ طفلاً. قد يستنتج أن الهدف كان إنقاذ الصبي وأن المبادئ الأخلاقية تجعل الإطفائي يقدر حياة الصبي أكثر من الاسترخاء في عربة الإطفاء - وفي الواقع يقدره بما يكفي للمخاطرة بسلامته الشخصية. ولكن الذكاء الاصطناعي قد يستنتج عوضا عن ذلك أن الإطفائي كان يتجمد من البرد ويشتاق للحرارة، أو أنه فعل ذلك كتمرين رياضي. إذا كان هذا المثال هو المثال الوحيد الذي يعرفه الذكاء الاصطناعي عن الإطفائيين والحريق والأطفال، فبالطبع سيكون من المستحيل معرفة أي تفسير كان صحيحا. ولكن الفكرة الأساسية الكامنة وراء التعلّم بالتعزيز العكسي هي أننا نتخذ القرارات طوال الوقت، وأن كل قرار نتخذه يكشف عن شيء ما حول أهدافنا. لذا، من المأمول أن يتمكن الذكاء الاصطناعي من بناء نموذج دقيق لجميع تفضيلاتنا من خلال مراقبة الكثير من الأشخاص في الكثير من المواقف (سواء أكانت حقيقية أم في الأفلام أم الكتب).[4]

في مقاربة التعلم التعكسي، تتمثل الفكرة الأساسية في أن الذكاء الاصطناعي يحاول تحقيق ليس أقصى قدر من تحقيق الهدف نفسه، بل من رضا مالكه البشري.

لذا، يكون لديه حافز لتوخي الحذر عندما يكون غير متأكد مما يريده مالكه، وبذل قصارى جهده فهمه.

كما يجب ألا ينزعج من قيام مالكه بإيقاف تشغيله، لأن ذلك قد يعني أنه قد أساء فهم ما يريده مالكه حقاً.

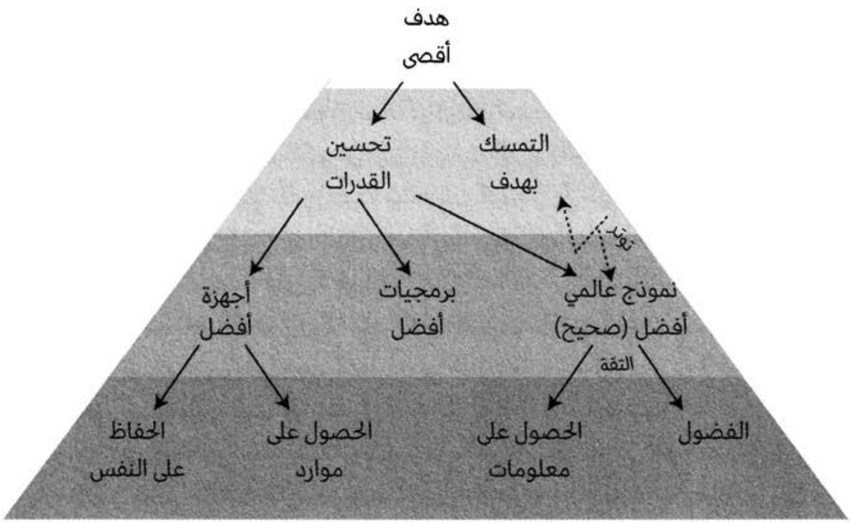
حتى لو كان من الممكن صنع ذكاء اصطناعي يتعلم أهدافَك، فإن هذا لا يعني أنه سيتبنّاها بالضرورة. انظر مثلا إلى السياسيين الأقل تفضيلا لديك: أنت تعرف ما يريدونه. لكن هذا ليس ما تريده، وعلى الرغم من أنهم يحاولون جاهدين، فقد فشلوا في إقناعك باعتماد أهدافهم.

لدينا العديد من الاستراتيجيات لتَشرُّب أطفالنا أهدافنا - وبعضها أكثر نجاحاً من غيرها، كما تعلمتُ من تنشئة صبيين مراهقين. وعندما يكون الطرف الذي يجب إقناعه عبارة عن أجهزة حاسوب بدلاً من أشخاص، فإن التحدي يُعرف باسم **مشكلة تحميل القيمة** ***Value-loading problem***، وهو أصعب من التعليم الأخلاقي للأطفال. خذ بعين الاعتبار نظام ذكاء اصطناعي يتم تحسينه تدريجيّاً من مستوى دون بشري إلى مستوى يفوق البشر، أولا من قبلنا نحن ثم من خلال التحسين الذاتي المتكرر مثل بروميثيوس. في البدء، يكون أقل قوة منك، لذلك لا يمكنه منعك من إيقاف تشغيله واستبدال تلك الأجزاء من برمجياته وبياناته التي تُرمِّز أهدافه - لكن هذا لن ينفعك، لأنه لا يزال من الغباء بما يمنعه من أن يفهم أهدافك فهما تاما، والتي تتطلب ذكاء على المستوى البشري لفهمها. وفي النهاية، سيكون أكثر ذكاءً منك، ومن المأمول أن يكون قادرا على فهم أهدافك فهما تاما - لكن هذا ربما لا ينفعك أيضاً، لأنه الآن أكثر قوة منك، وربما لا يسمح لك بإغلاقه واستبدال أهدافه مثلما لا تدع -أنت- أولئك السياسيين يستبدلون أهدافك بأهدافهم.

بعبارة أخرى، النافذة الزمنية التي يمكنك خلالها تحميل أهدافك في الذكاء الاصطناعي قد تكون قصيرة جدا: الفترة القصيرة بين الوقت الذي يكون فيه غبياً جدّاً كي يفهمك وذكياً جدّاً بحيث لا يسمح لك بإيقاف تشغيله. السبب في أن تحميل القيمة قد يكون أكثر صعوبة مع الأجهزة منه في البشر هو أن نمو ذكائها ربما يكون أسرع بكثير: فبنما قد يقضي الأطفال سنوات طويلة في تلك النافذة السحرية القابلة للإقناع حيث يكون ذكاءهم معادلا لذكاء آبائهم، فإن الذكاء الاصطناعي، مثل بروميثيوس، يندفع خارجا من هذه النافذة في غضون أيام أو ساعات.

ويتتبع بعض الباحثين مقاربة بديلة لجعل الآلات تتبنى أهدافنا، تعرف بالمصطلح الطنان القابلية للإصلاح Corrigibility. ومن المأمول أن يتمكن المرء من منح الذكاء الاصطناعي البدائي نظام أهداف، بحيث لا يهتم ببساطة إذا أوقفت تشغليه من وقت لآخر وغيّرت أهدافه. إذا كان هذا ممكناً، فيمكنك إذاً السماح للذكاء الاصطناعي بأن يصبح ذكاء فائقا، ومن ثم إيقاف تشغيله، وتثبيت Install أهدافك، وتجربتها لفترة من الوقت، والوقت الذي كنت فيه غير راضٍ عن النتائج، فقط أجرِ المزيد من التعديلات على الأهداف.

لكن حتى لو كنت تبني ذكاء اصطناعيا سيتعلم أهدافك ويتبناها، فإنك ما زلت لم تحل مشكلة مواءمة الأهداف: ماذا لو تطورت أهداف ذكائك الاصطناعي مع ازدياد ذكائه؟ كيف ستضمن أنه سيظل متمسكا بأهدافك بغض النظر عن مقدار التحسين الذاتي المتكرر الذي يمر به؟ دعونا نستكشف حجة مثيرة للاهتمام حول كون التمسك بالهدف مضمون تلقائياً، ومن ثم معرفة ما إذا كان بإمكاننا اكتشاف عيوب في تلك الحجة.على الرغم من أننا لا نستطيع التنبؤ بالتفصيل بما سيحدث بعد انفجار ذكاء - ولهذا السبب أطلق عليه فيرنون فينغ Vernor Vinge "حدثاً منفرداً" Singularity - جادل الفيزيائي والباحث في الذكاء الاصطناعي ستيف أوموهوندرو Steve Omohundro في مقال نشر في عام 2008 أنه يمكننا مع ذلك توقع بعض جوانب سلوك الذكاء الاصطناعي الفائق الذكاء بشكل مستقل تقريباً عن أي من أهدافه النهائية.[5] وقد تمت مراجعة هذه الحجة وتطويرها بشكل أكبر في كتاب نيك بوستروم **الذكاء الفائق** ***Superstar Intelligence***. الفكرة الأساسية هي أنه مهما كانت أهدافه النهائية، فستؤدي هذه إلى أهداف فرعية يمكن التنبؤ بها. فيما سبق من هذا الفصل، رأينا الكيفية التي أدى بها هدف النسخ إلى هدف فرعي هو الأكل، وهو ما يعني أنه على الرغم من ليس من الممكن لكائن فضائي يراقب بكتيريا الأرض الآخذة بالتطور منذ بلايين السنين أن يتنبأ بما ستكون عليه جميع أهدافنا البشرية، فإنه قد يتوقع بأن يكون أحد أهدافنا هو الحصول على المواد الغذائية. إذا نظرنا إلى المستقبل، ما الأهداف الفرعية التي ينبغي أن نتوقعها لذكاء اصطناعي فائق؟



الشكل 2.7: أي هدف نهائي لذكاء اصطناعي فائق يؤدي بطبيعة الحال إلى الأهداف الفرعية المبينة في الشكل. ولكن هناك توتر متأصل بين التمسّك بهدف وتحسين نموذجه عن العالم؛ مما يثير الشكوك حول ما إذا كان سيتمسك فعليّاً بالهدف الأصلي عندما يصبح أكثر ذكاءً.

ما أراه، هو أن الحجة الأساسية هي أنه لإتاحة أقصى قدر من الفرص لتحقيق أهدافه القصوى، أيا كانت، ينبغي للذكاء الاصطناعي متابعة الأهداف الفرعية الموضحة في الشكل 2.7. يجب أن يسعى ليس فقط إلى تحسين قدرته على تحقيق أهدافه القصوى، ولكن أيضاً لضمان أنه سيتمسك بهذه الأهداف حتى بعد أن صار أكثر قدرة. هذا يبدو مقبولاً تماماً: ففي المحصلة، هل ستختار الحصول على غرسة دماغية تُعزز الذكاء إذا علمت أن ذلك سيجعلك تريد قتل أحبائك؟ هذه الحجة القائلة إن ذكاءً اصطناعيا أكثر ذكاءً سيتمسك بأهدافه القصوى تُشكِّل حجر زاوية في رؤية الذكاء الاصطناعي الودود التي يروّج لها اليعازر يودكوفسكي وآخرون: وهي تقول بشكل أساسي إننا إذا تمكنا من تطوير ذكاء اصطناعي يُحسِّن ذاته ليصبح ودودا من خلال التعلم ويتبنى أهدافِنا، إذاً فإننا حققنا ما نريد، لأننا قد ضمنا أنه سيبذل قصارى جهده للبقاء ودوداً إلى الأبد.

ولكن هل هذا صحيح حقا؟ للإجابة عن هذا السؤال، نحتاج أيضاً إلى استكشاف الأهداف الفرعية الناشئة الأخرى في الشكل 2.7. من الواضح أن الذكاء الاصطناعي سوف يزيد من فرصته في تحقيق هدفه الأقصى، أيا كان، إذا كان بإمكانه تعزيز قدراته، ويمكنه القيام بذلك عن طريق تحسين أجهزته وبرمجياته* ونموذجه عن العالم. الأمر نفسه ينطبق علينا نحن البشر: الفتاة التي تهدف إلى أن تصبح أفضل لاعبة تنس في العالم سوف تتدرب على تحسين أجهزتها العضلية للعب التنس، وبرنامجها العصبي للتنس، ونموذجها العقلي عن العالم الذي يساعد على التنبؤ بما سيفعله خصومها. بالنسبة إلى الذكاء الاصطناعي، فإن الهدف الفرعي لأمثلة أجهزته يُفضِّل الاستخدام الأفضل للموارد الحالية (أجهزة الاستشعار والمحركات والحوسبة وما إلى ذلك) واكتساب المزيد من الموارد، كما يتضمن رغبة في الحفاظ على الذات، لأن التدمير/إيقاف التشغيل سيؤديان إلى التدهور النهائي للأجهزة.

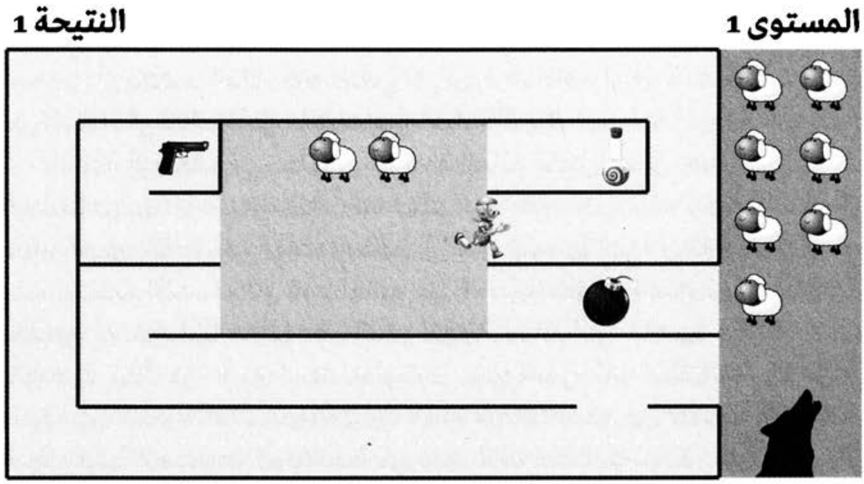
لكن انتظر لحظة! ألم نقع في فخ إسباغ السمات البشرية على الذكاء الاصطناعي بكل هذا الحديث عن الكيفية التي سيحاول بها جمع الموارد والدفاع عن نفسه؟ ألا ينبغي لنا أن نتوقع مثل هذه السمات النمطية للذكر السائد Alpha-male فقط في الذكاءات الناشئة عن التطور الداروني شرس التنافس؟ نظراً لأن الذكاء الاصطناعي مصمم هندسيا ولم يتطور عضويا، ألا يمكن أن تكون بالقدر نفسه من اللاطموح والتضحية بالنفس؟

* أنا أستخدم مصطلح «تحسين برمجياته» بأوسع معنى ممكن، بما في ذلك ليس فقط تحسين الخوارزميات الخاصة به ولكن أيضاً جعل عملية صنع القرار الخاصة به أكثر عقلانية، بحيث يصبح تحقيق أهدافه في متناول يديه قدر الإمكان.

كدراسة حالة بسيطة، دعونا نتأمل روبوت الذكاء الاصطناعي في الشكل 3.7. وهدفه الوحيد هو إنقاذ أكبر عدد ممكن من الأغنام من الذئب الشرير الكبير. هذا يبدو وكأنه هدف نبيل وإيثار لا علاقة له تماماً بالحفاظ على الذات واكتساب الموارد. ولكن ما هي أفضل استراتيجية لصديقنا الروبوت؟ فلن ينقذ الروبوت المزيد من الأغنام إذا صادف قنبلة، لذلك لديه حافز تجنب الانفجار بها. وبعبارة أخرى، فإنه يطور مجموعة فرعية من أهداف الحفاظ على الذات! كما أن لديه حافزاً لإظهار الفضول، وتحسين نموذجه عن العالم من خلال استكشاف بيئته، لأن المسار الذي يسير عليه حالياً سيصل به في النهاية إلى المراعي، غير أن هناك بديلاً أقصر سيتيح للذئب وقتاً أقل لافتراس الخراف. أخيراً، إذا اسكتشف الروبوت جميع الجوانب بتفصيل، فسيكتشف قيمة الحصول على الموارد: جرعة تجعله يعمل بشكل أسرع ومسدس يسمح له بإطلاق النار على الذئب. باختصار، لا يمكننا استبعاد الأهداف الفرعية "للألفا الذكورية" مثل الحفاظ على الذات واكتساب الموارد باعتبارها ذات صلة بالكائنات الحية المتطورة تطورا عضويا فقط، لأن روبوت الذكاء الاصطناعي الذي صنعناه طورها من هدفه الوحيد وهو تحقيق رفاهية الأغنام.

إذا شرّبت الذكاء الاصطناعي الفائق بهدف وحيد هو التدمير الذاتي، فسيفعل ذلك بسعادة. ولكن النقطة المهمة هي أنه سيقاوم إيقاف التشغيل إذا أعطيته أي هدف يحتاج إلى أن يظل يعمل لتحقيقه - وهذا يشمل جميع الأهداف تقريباً! إذا أعطيت للذكاء الفائق الهدف الوحيد المتمثل في تقليل الضرر الذي يلحق بالإنسانية، مثلاً، فسيدافع عن نفسه ضد محاولات إيقاف التشغيل لأنه يعلم أننا سنضر بعضنا البعض أكثر بكثير في غيابه بإشعال حروب مستقبلية وغيرها من الحماقات الأخرى.

وبالمثل، يمكن تحقيق جميع الأهداف تقريباً بشكل أفضل باستخدام المزيد من الموارد، لذلك يجب أن نتوقع من الذكاء الفائق أن يرغب في الحصول على موارد بغض النظر تقريباً عن هدفه الأقصى. ومن ثم، فمن الخطر إعطاء ذكاء اصطناعي فائق هدفا واحدا لانهائيا دون أي قيود: إذا أنشأنا ذكاءً فائقاً وهدفه الوحيد هو أن يلعب لعبة Go قدر الإمكان، فإن الشيء المنطقي للقيام بذلك هو إعادة ترتيب مجموعتنا الشمسية لتشكّل حاسوبا عملاقا دون أي اعتبار لسكانها السابقين، يبدأ باستيطان عالمنا للسعي للحصول على مزيد من القوة الحوسبية. وهكذا استكملنا الآن الدائرة: مثلما يعطي هدفُ الموارد بعضَ البشر هدفا فرعيا لإتقان اللعبة Go، فإن هذا الهدف من إتقان اللعبة Go يمكن أن يؤدي إلى هدف فرعي يدفع إلى الحصول على الموارد. في الخلاصة، هذه الدرجات الفرعية الناشئة تحتم ألا نطلق العنان للذكاء الفائق قبل حل مشكلة مواءمة الأهداف: ما لم نهتم اهتماماً كبيراً في منحها أهدافاً صديقة للإنسان، فقد تنتهي الأمور نهاية سيئة تحيق بنا.



الشكل 3.7: حتى لو كان الهدف النهائي للروبوت هو زيادة نتيجة إعادة الأغنام من المراعي إلى الحظيرة قبل أن يأكلها الذئب إلى أقصى حد، فإن هذا قد يؤدي إلى أهداف فرعية من الحفاظ على الذات (تجنب القنبلة)، الاستكشاف (البحث عن طريق مختصر) والحصول على الموارد (ترياق يجعله يعمل بشكل أسرع ومسدس يسمح له بإطلاق النار على الذئب).

نحن الآن على استعداد لمعالجة الجانب الثالث والشائك أكثر من الجانبين السابقين من مشكلة مواءمة الأهداف: إذا نجحنا في الحصول على ذكاء فائق **يتحسن** ذاتياً **ويتعلم** أهدافنا **ويتبناها**، فهل سيتمسك بها، كما جادل أوموهوندرو؟ ما الدليل؟

يزداد ذكاء البشر زيادة كبيرة أثناء نموهم. لكنهم لا يتمسكون دائماً بأهداف طفولتهم. على النقيض من ذلك، غالباً ما يغير الناس أهدافهم بشكل كبير عندما يتعلمون أشياء جديدة ويزدادون حكمة. كم عدد البالغين الذين تعرف ممن تُحفزهم مشاهدة المسلسل الكارتوني **تيليتبيز** ***Teletubbies***؟ لا يوجد أي دليل على أن تطور مثل هذا الهدف يتوقف عندما يتخطى الذكاء حداً معيناً -في الواقع، هناك تلميحات إلى أن الميل إلى تغيير الأهداف استجابةً لخبرات ورؤى جديدة يزداد بدلاً من أن يتناقص مع تزايد الذكاء.

لماذا الأمر على ما هو عليه؟ انظر مجددا في الأهداف الفرعية المذكورة آنفا لبناء نموذج أفضل للعالم من حولك -هنا تكمن المشكلة! هناك توتر بين نمذجة العالم والتمسك بالهدف (انظر: الشكل 2.7). فمع زيادة الذكاء ربما لا يأتي فقط تحسن كمي في القدرة على تحقيق الأهداف القديمة نفسها، ولكن فهما نوعي مختلف لطبيعة الواقع الذي يكشف أن الأهداف القديمة كانت مُضلِّلة،أو حتى غير مُحدَّدة. مثلاً، لنفترض أننا برمجنا ذكاء اصطناعيا ودودا لزيادة أعداد البشر الذين تذهب أرواحهم إلى الجنة في الحياة الآخرة. أولاً. سيحاول أمورا مثل زيادة تعاطف الناس مع بعضهم وزيادة ذهابهم إلى الكنيسة. لكن لنفترض أنه بعد ذلك يُشكل فهما مجردا خاصا به للإنسان والوعي

الإنساني، ويقرر ذاتيا أنه لا يوجد شيء اسمه الروح. ماذا سيفعل حينها؟ بالمثل، فمن المحتمل أن يكتشف الذكاء الاصطناعي -في نهاية المطاف- أن أي هدف نُسنده له بناء على فهمنا الحالي للعالم (مثل «زيادة مغزى الحياة البشرية») هو هدف غير مُحدّد بوضوح.

إضافة إلى ذلك، في محاولاته لنمذجة العالم بشكل أفضل، قد يحاول الذكاء الاصطناعي بشكل طبيعي، تماماً كما فعلنا البشر، أن يفهم نفسه وكيفية عمله -أي بعبارة أخرى، التأمل في الذات. وبمجرد أن يبني نموذجاً جيداً لذاته ويفهم ما هو عليه، سيفهم الأهداف التي قدمناها له على مستوى الخطاب التعريفي، وربما يختار تجاهلها أو تقويضها، مثلما نحن البشر نفهم الأهداف التي حدّدتها جيناتنا ونتعمد تقويضها، مثلا عن طريق استخدام وسائل منع الحمل. لقد استكشفنا بالفعل في قسم علم النفس آنفا سبب اختيارنا لخداع جيناتنا وتقويض هدفها: لأننا نشعر بالولاء فقط للمجموعة المتناقضة من تفضيلاتنا العاطفية، وليس للهدف الوراثي الذي حفزها - وهو ما نفهمه الآن ونجده هدفا عاديا. لذلك اخترنا اختراق آلية المكافآت لدينا مستغلين ثغراتها. وبالمثل، فإن هدف حماية القيمة البشرية التي نبرمجها في الذكاء الاصطناعي الودود يصبح جينات الآلة، وبمجرد أن يفهم الذكاء الاصطناعي الودود نفسه جيداً بما فيه الكفاية، فإنه قد يعتبر هذا الهدف عادياً كما نشعر حيال التكاثر الإجباري، وليس من الواضح أنه لن يجد طريقة لتقويض ذلك مستغلا الثغرات الموجودة في برمجياتنا.

مثلا، لنفترض أن مجموعة من النمل صنعت روبوتاً يتحسن ذاتياً بشكل مطّرد، وأكثر ذكاءً منها، ويشاركها أهدافها ويساعدها على بناء تلال جحور نمل أكبر وأفضل، وأنه في النهاية سيصل إلى الذكاء على المستوى البشري ويفهم أنه يمتلك ذلك. هل تعتقد أنه سيقضي بقية أيامه فقط في تحسين تلال جحور النمل، أم هل تعتقد أنه قد يُطور ذوقاً لأسئلة ومساعٍ أكثر تطوراً لا يستطيع النمل فهمها؟ إذا كان الأمر كذلك، هل تعتقد أنه سيجد طريقة تُبطل Override بها برمجة نداء حماية النمل الذي برمجه صانعوه من النمل بالطريقة نفسها التي أبطلت أنت بها بعض الرغبات التي تُمليها جيناتك؟ وفي هذه الحالة، هل يمكن لذكاء فائق ودود أن يجد أهدافنا الإنسانية الحالية أهدافا غير ملهمة وباعثة على الضجر مثل أهداف النمل، فيطور أهدافاً جديدة مختلفة عن تلك التي تعلمها وتبناها منا؟

ربما هناك طريقة لتصميم ذكاء اصطناعي يتحسن ذاتياً ويضمن الاحتفاظ بالأهداف الصديقة للإنسان إلى الأبد. لكنني أعتقد أنه من العدل أن نقول إننا لا نعرف بعد كيف نبني واحداً - أو حتى ما إذا كان ذلك ممكناً. في الختام، تشتمل مشكلة مواءمة أهداف الذكاء الاصطناعي على ثلاثة أجزاء، لم يتم حل أي منها، وكلها الآن موضوع بحث جاد. ونظراً لأنها مشكلات عويصة جدّاً، فمن الأكثر أماناً البدء في تكريس أفضل جهودنا لحلها الآن، قبل وقت طويل من تطوير أي تقنية ذكية فائقة، لضمان وجود الإجابات عندما نحتاج إليها.

أخلاق: اختيار الأهداف

لقد استكشفنا الآن كيفية جعل الآلات تتعلم أهدافنا وتبَناها وتتمسك بها. لكن من "نحن"؟ من الذي نتحدث عنه؟ هل يجب أن يقرر شخص أو مجموعة الأهداف التي يتبناها الذكاء الفائق في المستقبل، على الرغم من الاختلاف كبير بين أهداف أدولف هتلر والبابا فرانسيس وكارل ساغان؟ أم أن هناك نوعا من الأهداف المتفق عليها والتي تشكل حلا وسطا جيدا للإنسانية ككل؟

في رأيي، إن هذه المشكلة الأخلاقية ومشكلة مواءمة الأهداف هي مشكلات جوهرية تحتاج إلى حل قبل تطوير أي ذكاء فائق. فمن ناحية، فإن تأجيل العمل على القضايا الأخلاقية إلى ما بعد بناء الذكاء الفائق الموائم لأهدافنا سيكون قرارا غير مسؤول وربما كارثيا. إن ذكاء فائقا مطيعا تماماً تتواءم أهدافه تلقائياً مع أهداف مالكه البشري سيكون مثل نسخة مضاعفة من أدولف أيخمان Adolf Eichmann كرئيس مكتب الأمن الرئيسي للرايخ الثالث النازي: يفتقر إلى البوصلة الأخلاقية أو موانع ناهية ذاتية. ومهما كانت تلك الأهداف فإنه سينفذ أهدافه بفعالية شرسة.[6] من ناحية أخرى، فقط إذا نجحنا في حل مشكلة محاذاة الأهداف فقط، فيتاح لنا رفاهية الجدال حول الأهداف التي يجب تحديدها. الآن لننغمس في هذه الكماليات.

منذ العصور القديمة، حلم الفلاسفة في اشتقاق الأخلاق Ethics (المبادئ التي تحكم كيفية تصرفنا) من الأساس فقط باستخدام المنطق ومبادئ غير قابلة للجدل. للأسف، بعد آلاف السنين، فإن الإجماع الوحيد الذي تم التوصل إليه هو أنه لا يوجد أي إجماع. مثلاً، بينما أكد أرسطو على الفضائل Virtues، أكد إيمانويل كانت على الواجبات Duties، وشدد النفعيون على أعظم سعادة تكمن في أكبر عدد ممكن. وجادل كانت بأنه قادر على استنباط استنتاجات من المبادئ الأولى (التي أطلق عليها "الضرورات الفئوية" Categorical imperatives) وهي استنتاجات قد يعارضها العديد من الفلاسفة الغربيين المعاصرين، مثل: العادة السرية أسوأ من الانتحار، الشذوذ الجنسي أمر بغيض، من الجيد قتل الأوغاد، الزوجات والخدم والأطفال مملوكون مثل المتاع.

من ناحية أخرى، على الرغم من هذا الخلاف، هناك العديد من الموضوعات الأخلاقية التي يوجد حولها اتفاق واسع النطاق، عبر الثقافات وعبر القرون. مثلاً، التركيز على **الحق والجمال والخير** تعود إلى البهاغافاد غيتا Bhagavad Gita وأفلاطون. فشعار معهد الدراسات المتقدمة Institute for Advanced Study في جامعة برينستون، حيث عملت سابقاً في مرحلة ما بعد الدكتوراه، هو "الحقيقة والجمال" Truth & Beauty، في حين أهملت جامعة هارفارد الجمال وأكّدت ببساطة على "الحقيقة" Veritas. في كتابه **سؤال جميل** ***A Beautiful Question***، يجادل زميلي فرانك ويلكزِك Frank Wilczek في أن الحقيقة مرتبطة بالجمال ويمكننا أن ننظر إلى عالمنا كعمل فني. أما العلم والأديان والفلسفة فكلها تطمح إلى الحقيقة. وتركز الأديان بقوة على الخير، وكذلك فعلت جامعتي،

المعهد MIT: في خطابه الافتتاحي في عام 2015، أكد رئيسنا، رافائيل ريف Rafael Reif، على مهمتنا لجعل عالمنا مكاناً أفضل.

على الرغم من أن محاولات التوصل إلى إجماع حول الأخلاق قد فشلت حتى الآن، فإن هناك اتفاقا واسعا على أن بعض المبادئ الأخلاقية نابعة من مبادئ أكثر أساسية، مثلما تنبع الأهداف الفرعية من الأهداف الأساسية. مثلاً، يمكن النظر إلى الطموح للوصول إلى الحقيقة باعتباره البحث عن نموذج أفضل للعالم كما هو موضح في الشكل 2.7: فهم الطبيعة المطلقة للواقع يساعد على تحقيق أهداف أخلاقية أخرى. بالطبع، لدينا الآن إطار ممتاز للبحث عن الحقيقة: المنهج العلمي Scientific method. ولكن كيف يمكننا تحديد ما هو جميل أو خيّر؟ ويمكن أيضاً أن تعزى بعض جوانب الجمال إلى الأهداف الأساسية، مثلاً، قد تؤثر معاييرنا الخاصة بجمال الذكور والإناث جزئياً في تقييمنا اللاوعي لملاءمة تكرار جيناتنا.

فيما يتعلق بالخير، فإن ما يسمى بالقاعدة الذهبية Golden Rule (ينبغي للمرء أن يعامل الآخرين كما يود أن يعاملوه) في معظم الثقافات والأديان، ويهدف بوضوح إلى تشجيع الاستمرار المتناغم للمجتمع البشري (ومن ثم جيناتنا) من خلال تعزيز التعاون وتثبيط الفتنة غير المنتجة.[7] ويمكن قول الشيء نفسه عن العديد من القواعد الأخلاقية المُكرّسة في النظم القانونية في جميع أنحاء العالم، مثل التركيز الكونفوشيوسي على الصدق، والعديد من الوصايا العشر Ten Commandments، بما في ذلك «لا تقتل». بعبارة أخرى، هناك العديد من العناصر الأخلاقية التي تشترك مع المشاعر الاجتماعية مثل التعاطف والرحمة: لقد تطورت لتوليد التعاون، وتؤثر في سلوكنا من خلال المكافآت والعقوبات، إذا قمنا بشيء ما وشعرنا بالسوء بعد ذلك، فذلك عقابنا العاطفي الناجم مباشرة من كيمياءُ الدماغ. من ناحية أخرى، إذا انتهكنا المبادئ الأخلاقية، فقد يعاقبنا المجتمع بشكل غير مباشر، مثلا بالخزي غير الرسمي بفعل أقراننا أو معاقبتنا لخرق القانون.

بعبارة أخرى، على الرغم من أن الإنسانية اليوم ليست قريبة من الإجماع الأخلاقي، فإن هناك العديد من المبادئ الأساسية التي يوجد حولها اتفاق واسع. هذا الاتفاق ليس مفاجئاً، لأن المجتمعات البشرية التي وصلت إلى الوقت الحاضر تميل إلى أن تكون لها مبادئ أخلاقية مُؤَمثَلَة للهدف نفسه: الترويج لنجاتها ومن ثم ازدهارها. ونحن نتطلع إلى المستقبل حيث للحياة القدرة على الازدهار في جميع أنحاء عالمنا لبلايين السنين، ما هي أقل مجموعة من المبادئ الأخلاقية التي قد نتفق على أننا نريد لهذا المستقبل أن يفي بها؟ هذه محادثة نحتاج جميعا إلى أن نكون جزءاً منها. لقد كان من الرائع بالنسبة إلي أن أسمع وأقرأ وجهات النظر الأخلاقية للعديد من المفكرين على مدار سنوات عديدة، وحسب رأيي، يمكن تقسيم معظم تفضيلاتهم إلى أربعة مبادئ:

- النفعية Utilitarianism: يجب زيادة الخبرات الإيجابية الواعية إلى أقصى حد وينبغي الحد من المعاناة.

- التنوع Diversity: مجموعة متنوعة من التجارب الإيجابية أفضل من العديد من التكرارات للتجربة نفسها، حتى لو اعتبرت فيما بعدها بأنها أفضل تجربة إيجابية ممكنة.
- الاستقلالية Autonomy: يجب أن تتمتع الكيانات/المجتمعات الواعية بحرية السعي لتحقيق أهدافها ما لم يتعارض هذا مع مبدأ سائد.
- الإرث Legacy: التوافق مع السيناريوهات التي يجدها معظم البشر اليوم سيناريوهات سعيدة، وعدم التوافق مع السيناريوهات التي يعتبرها جميع البشر اليوم رهيبة.

لنأخذ لحظة لتحليل واستكشاف هذه المبادئ الأربعة. تقليديا، يُنظر إلى النفعية على أنها تعني «أعظم سعادة لأكبر عدد من الأشخاص». لكني عمّمت الفكرة هنا لتكون أقل تمركزا حول الإنسان، بحيث يمكن أن تشمل أيضاً الحيوانات غير البشرية، ومحاكاة العقول البشرية الواعية، وغيرها من أشكال الذكاء الاصطناعي الذي قد يوجد في المستقبل. وحدّدت التعريف من حيث **التجارب** بدلاً من الأشخاص أو الأشياء، لأن معظم المفكرين يتفقون على أن الجمال والفرح والسرور والمعاناة هي تجارب ذاتية. وهذا يعني أنه إذا لم تكن هناك تجربة (كما هي الحال في عالم ميت أو مأهول بآلات لا وعي لها تشبه الزومبي)، فلن يكون هناك معنى أو أي شيء آخر ذي صلة أخلاقيا. إذا وافقنا على هذا المبدأ الأخلاقي النفعي، فمن الضروري أن نحدد أيّاً من الأنظمة الذكية واعيةً (بمعنى أن يكون لها تجربة ذاتية) وأيّاً منها ليس واعيا؛ وهذا هو موضوع الفصل التالي.

إذا كان هذا المبدأ النفعي هو المبدأ الوحيد الذي نهتم به، فعندئذ قد نرغب في معرفة ما هي أكثر تجربة إيجابية ممكنة، ثم توطين عالَمنا وإعادة إنتاج هذه التجربة نفسها بالضبط (وليس أي شيء آخر) مراراً وتكراراً مرة أخرى، أكبر عدد ممكن من المرات في أكبر عدد ممكن من المجرات - باستخدام عمليات المحاكاة إذا كانت هذه هي الطريقة الأكثر فعالية. إذا كنت تشعر أن هذه طريقة عادية جدّاً لاستغلال وقفنا الكوزموسي، فأظن أن جزءاً على الأقل مما تعتقد أن هذا السيناريو يفتقر إليه هو التنوع. كيف سيكون شعورك إذا كانت جميع وجباتك لبقية حياتك متطابقة؟ إذا كانت جميع الأفلام التي شاهدتها على الإطلاق هي الأفلام نفسها؟ إذا بدا كل أصدقائك متطابقين وكان لديهم شخصيات وأفكار متطابقة؟ ربما ينبع جزء من تفضيلنا للتنوع من كونه ساعد البشرية على البقاء والازدهار، بجعلنا أكثر صلادة. ربما يرتبط أيضاً بتفضيل الذكاء: فتنامي الذكاء خلال 13.8 بليون سنة من تاريخنا الكوني حوَّل الوَحدة المُملة إلى بُنى أكثر تنوعاً وتمايزاً وتعقيدا تُعالج المعلومات بطرق تزداد تعقيدا.

مبدأ الاستقلالية هو أساس العديد من الحريات والحقوق المُحدَّدة في الإعلان العالمي لحقوق الإنسان الذي اعتمدته الأمم المتحدة في عام 1948 في محاولة للتعلم من دروس الحربين العالميتين. وهذا يشمل حرية الفكر والكلام والحركة، والتحرر من العبودية

والتعذيب، والحق في الحياة والحرية والأمن والتعليم والحق في الزواج والعمل والملكية الخاصة. وإذا كنا نرغب في أن نكون أقل تركيزا على السمات البشرية، فيمكننا تعميم ذلك على حرية التفكير والتعلم والتواصل والملكية الخاصة وعدم التعرض للأذى، والحق في فعل أي شيء لا ينتهك حريات الآخرين. يساعد مبدأ الاستقلالية على التنوع، مادام الجميع لا يتشاركون في الأهداف نفسها بالضبط. إضافة إلى ذلك، فإن مبدأ الاستقلالية هذا ينبع من مبدأ المنفعة إذا كانت للكيانات الفردية تجارب إيجابية كأهداف وتحاول التصرف بما يحقق مصلحتها الذاتية: إذا كنا بدلاً من ذلك نمنع الكيان من متابعة تحقيق هدفه على الرغم من أن هذا لن يسبب أي ضرر لأي شخص آخر، هناك فستكون هناك خبرات إيجابية أقل عموما. بالطبع، فإن حجة الاستقلالية هذه هي بالضبط الحجة التي يستخدمها الاقتصاديون لصالح سوق حرة Free market: إنها تؤدي بطبيعة الحال إلى وضع فعال Efficient (يدعوها الاقتصاديون " أمثلية باريتو" Pareto-optimality) إذ لا يمكن لأحد أن يتحسن حاله دون أن تسوء حال شخص آخر.

يقول مبدأ الإرث بشكل أساسي إنه يجب أن يكون لدينا رأي في المستقبل لأننا نساعد على إنشائه. إن كلا من مبدأي الاستقلالية والإرث يجسدان المثل العليا الديمقراطية: فالأول يعطي أشكال الحياة في المستقبل سلطة على كيفية استخدام الوقف الكوزموسي، في حين أن الآخر يمنح البشر اليوم بعض القوة على ذلك.

على الرغم من أن هذه المبادئ الأربعة قد تبدو غير مثيرة للجدل، فإن تطبيقها في الممارسة العملية أمر صعب لأن المشكلات تكمن في التفاصيل. المشكلة تشبه معضلات «قوانين الروبوتات الثلاثة» التي وضعها مؤلف الخيال العلمي آيزاك أسيموف:

1. لا يصيب الروبوت أي إنسان بأي إصابة من خلال التقاعس عن العمل، أو السماح بأن يقع ضرر على الإنسان.
2. يجب أن يطيع الروبوت الأوامر التي أعطاه إياها إنسان إلا عندما تتعارض هذه الأوامر مع القانون الأول.
3. يجب أن يحمي الروبوت وجوده الذاتي مادامت هذه الحماية لا تتعارض مع القانون الأول أو الثاني.

على الرغم من أن كل هذا يبدو جيداً، إلا أن العديد من قصص أسيموف تُظهر الكيفية التي تؤدي بها القوانين إلى تناقضات إشكالية في مواقف غير متوقعة، لنفترض الآن أننا نستبدل هذه القوانين بقانونين فقط، في محاولة لترميز الاستقلالية لأشكال الحياة المستقبلية:

1. للكيان الواعي حرية التفكير والتعلم والاتصال والملكية الخاصة وعدم التعرض للأذى أو التدمير.
2. للكيان الواعي الحق في فعل أي شيء لا يتعارض مع القانون الأول.

يبدوان قانونين جيدين، أليس كذلك؟ ولكن يرجى التفكير في هذا للحظة. إذا كانت الحيوانات واعية، فما الذي يفترض أن تأكله الحيوانات المفترسة؟ هل يجب أن يصبح جميع أصدقائك نباتيين؟ إذا كانت بعض برمجيات الحاسوب المستقبلية المعقدة تتحول إلى وعي، فهل من غير القانوني إنهاء هذه البرمجيات؟ إذا كانت هناك قواعد ضد إنهاء أشكال الحياة الرقمية، فعندئذ تحتاج أيضاً إلى فرض قيود على إنشائها لتجنب حدوث انفجار سكان رقمي؟ كان هناك اتفاق واسع النطاق على الإعلان العالمي لحقوق الإنسان فقط لأننا لم نسأل سوى البشر. بمجرد أن نأخذ بعين الاعتبار مجموعة أوسع من الكيانات الواعية بدرجات متفاوتة من القدرة والقوة، فإننا نواجه مقايضات صعبة بين حماية الضعفاء ومبدأ «القوة تفرض ما هو الصواب».

كذلك هناك مشكلات شائكة في مبدأ الإرث. نظرا إلى كيفية تطور وجهات النظر الأخلاقية منذ العصور الوسطى فيما يتعلق بالعبودية وحقوق المرأة وما إلى ذلك، هل نريد حقّاً أن يكون للناس منذ 1,500 عام تأثير كبير في كيفية إدارة عالَم اليوم؟ إذا كان الجواب بالنفي، فلماذا نحاول فرض أخلاقنا على كائنات مستقبلية قد تكون أكثر ذكاءً منا؟ هل نحن واثقون حقّاً من أن ذكاء فائقا AGI يريد ما تُثمِّنه عقولنا الأدنى ذكاء؟ سيكون هذا بمثابة طفلة عمرها أربع سنوات تتخيل أنها بمجرد أن تكبر وتصبح أكثر ذكاءً، فإنها سترغب ببناء منزل كبير من كعكة الزنجبيل حيث يمكن أن تقضي طوال اليوم في تناول الحلوى والآيس كريم. ومثلها، من المحتمل أن تتخطى الحياة على الأرض اهتمامات طفولتها. أو تخيل فأراً يصنع ذكاء فائقا AGI على المستوى البشري، وسيود إنشاء مدن بكاملها من الجبن. من ناحية أخرى، إذا علمنا أن ذكاء اصطناعيا خارقا سوف يرتكب يوماً ما جريمة إبادة الكون وإبادةجميع أشكال الحياة في عالمنا، فلماذا يجب أن يوافق بشر اليوم على هذا المستقبل الخالي من الحياة إذا كانت لدينا القدرة على منعه بإنشاء ذكاء اصطناعي للغد بطريقة مختلفة؟

في الختام، من الصعب ترميز حتى مبادئ الأخلاقيات المقبولة على نطاق واسع في شكل قابل للتطبيق على الذكاء الاصطناعي المستقبلي، وهذه المشكلة تستحق مناقشة وبحثا جديين مع استمرار الذكاء الاصطناعي في التقدم. ولكن في غضون ذلك، دعونا لا نتحول إلى عدو تام للخير: فهناك العديد من الأمثلة على "أخلاقيات رياض الأطفال Kindergarten ethics" التي لا جدال فيها، والتي يمكن بل ويجب دمجها في تكنولوجيا الغد. مثلاً، يجب ألا يُسمح لطائرات ركاب مدنية كبيرة بالطيران مخترقة أجساما ثابتة، والآن بعد أن صار جميعها تقريباً مزودا بالطيار الآلي والرادار ونظام تحديد المواقع العالمي GPS، لم يعد هناك أي أعذار تقنية. ومع ذلك، اخترق مختطفو الطائرات في 11 سبتمبر المباني بثلاث طائرات، وقاد الطيار الانتحاري أندرياس لوبيتز Andreas Lubitz الرحلة رقم 9525 التابعة للخطوط الجوية جيرمان وينغز Germanwings لتصطدم بجبل في 24 مارس 2015 - بوضع الطيار الآلي ليطير على ارتفاع 100 قدم (30 متراً) فوق مستوى سطح البحر والسماح لجهاز كمبيوتر الرحلة بالقيام بباقي العمل. الآن ومع تزايد ذكاء أجهزتنا بما يكفي للحصول على بعض المعلومات حول ما

تقوم به، فقد حان الوقت لتعليمها الحدود. يحتاج أي مهندس يصمم جهازاً إلى التساؤل عما إذا كانت هناك أشياء يمكنها القيام به ولكن يجب ألا تقوم بها، والنظر فيما إذا كانت هناك طريقة عملية تجعل من المستحيل على مستخدم شرير أو أخرق التسبب في ضرر.

أهداف قصوى؟

كان هذا الفصل تاريخاً موجزاً للأهداف. إذا أمكننا مشاهدة إعادة عرض سريعة لتاريخنا الكوني الذي يبلغ 13.8 بليون عام، فسنشهد عدة مراحل متميزة من السلوك الموجّه نحو الأهداف:

1. المادة على ما يبدو عازمة على تحقيق أقصى قدر لتبديد نفسها
2. الحياة البدائية تحاول على ما يبدو تحقيق أقصى قدر من تكرار نفسها
3. البشر لا يسعون إلى النَسخ ولكن إلى تحقيق أهداف تتعلق بالسعادة والفضول والتعاطف والمشاعر الأخرى التي تطورت لمساعدتهم على النسخ
4. آلات بنيت لمساعدة البشر على السعي إلى تحقيق أهداف البشرية

إذا أطلقت هذه الآلات في نهاية المطاف انفجار ذكاء، فكيف *سينتهي* تاريخ الأهداف في النهاية؟ هل يمكن أن يكون هناك نظام هدفي أو إطار عمل أخلاقي تتفق عليه جميع الكيانات تقريباً نظراً لأنها صارت أكثر ذكاءً؟ بعبارة أخرى، هل لدينا مصير أخلاقي من نوع ما؟

قد تشير القراءة السريعة للتاريخ البشري إلى تلميحات لمثل هذا التقارب: في كتابه **الملائكة الأفضل لطبيعتنا** ***The Better Angels of Our Nature***، يجادل ستيفن بينكر Steven Pinker بأن الإنسانية صارت أقل عنفاً وأكثر تعاوناً منذ آلاف السنين، وأن العديد من أنحاء عالمنا شهدت زيادة تقبل التنوع والاستقلالية والديمقراطية. إشارة أخرى على التقارب هو أن السعي وراء الحقيقة باستخدام المنهج العلمي اكتسب شعبية على مدى آلاف السنين الماضية. ولكن، هذه التوجهات قد تظهر ليس تقارباً بين الأهداف القصوى ولكن مجرد الأهداف الفرعية. مثلاً، يوضح الشكل 2.7 أن السعي وراء الحقيقة (نموذج للعالم أكثر دقة) هو ببساطة هدف فرعي تقريباً لأي هدف أقصى. وبالمثل، فقد رأينا آنفا كيف يمكن النظر إلى المبادئ الأخلاقية مثل التعاون والتنوع والاستقلال كأهداف فرعية، من حيث إنها تساعد المجتمعات على العمل بفعالية، ومن ثم مساعدتهم على البقاء على قيد الحياة وتحقيق أي أهداف أكثر جوهرية قد تكون لديهم. بل إن البعض قد يرفض كل شيء نسميه "القيم الإنسانية" Human values على أنه ليس سوى بروتوكول تعاون، يساعدنا على تنفيذ الهدف الفرعي للتعاون بشكل أكثر فاعلية. وبنفس الروح، ومتطلعين نحو المستقبل، من المحتمل أن يكون لأي ذكاء اصطناعي فائق الذكاء أهداف فرعية بما في ذلك الأجهزة الفعالة والبرمجيات الفعالة والبحث عن الحقيقة والفضول، وذلك ببساطة لأن هذه الأهداف الفرعية تساعدها على تحقيق أهدافها القصوى.

في الواقع، يجادل نيك بوستروم بقوة ضد فرضية المصير الأخلاقي في كتابه **الذكاء الفائق** ***Superintelligence***، إذ يقدم وجهة نظر مضادة يطلق عليها **فرضية التعامد** ***Orthogonality thesis***: إن الأهداف القصوى لأي نظام قد تكون مستقلة عن ذكائها. تعريفا، فإن الذكاء هو ببساطة القدرة على تحقيق أهداف معقدة، بغض النظر عن ماهية هذه الأهداف، ومن ثم فإن فرضية التعامد تبدو معقولة تماماً. ففي المحصلة، يمكن للناس أن يكونوا أذكياء ولطفاء أو أذكياء وقاسين، ويمكن استخدام الذكاء بهدف تحقيق اكتشافات علمية، أو إنشاء فنون جميلة، أو مساعدة الناس، أو التخطيط لهجمات إرهابية.[8]

فرضية التعامد تمكّننا من خلال إخبارنا بأن الأهداف القصوى للحياة في عالمنا ليست مقدرة سلفا، ولكن لدينا الحرية والقوة في تشكيلها. فهي تقترح أن التقارب المضمون نحو هدف فريد لا يمكن العثور عليه في المستقبل ولكن في الماضي، عندما برزت الحياة كلها بهدف واحد هو النسخ. ومع مرور الوقت الكوني، تحصل العقول الأكثر ذكاءً على فرصة للتمرد والتحرر من هذا الهدف المكرر للتكاثر واختيار أهداف خاصة بها. نحن البشر لسنا أحراراً تماماً بهذا المعنى، نظراً لأن العديد من الأهداف لا تزال مُكرّسة جدّاً في جيناتنا، ولكن يمكن أن يتمتع ذكاء اصطناعي بهذه الحرية المطلقة المتمثلة في عدم التقييد الكامل بالأهداف السابقة. هذا الاحتمال بنشوء تحرر أكثر من الهدف هو واضح في أنظمة الذكاء الاصطناعي الضيقة والمحدودة اليوم: كما ذكرت سابقاً، فإن الهدف الوحيد لحاسوب الشطرنج هو الفوز في لعبة الشطرنج، ولكن هناك أيضاً أجهزة حاسوب هدفها الخسارة في لعبة الشطرنج التي تتنافس في بطولات الشطرنج العكسية، إذ يكون الهدف هو إجبار الخصم على اقتناص قطعك. ولعل هذه التحرر من التحيزات التطورية يمكن أن يجعل الذكاء الاصطناعي أكثر أخلاقية من البشر بالمعنى العميق: جادل الفلاسفة الأخلاقيون مثل بيتر سينغر Peter Singer بأن معظم البشر يتصرفون بطريقة غير أخلاقية لأسباب تطورية، مثلا عن طريق التمييز ضد الحيوانات غير البشرية.

رأينا أن حجر الزاوية في رؤية "الذكاء الودود" هو فكرة مفادها أن الذكاء الاصطناعي الذي يتحسن بشكل مطّرد سوف يرغب في الحفاظ على هدفه الأقصى (الودي) لأنه يصبح أكثر ذكاءً، ولكن كيف يمكن لـ "هدف أقصى" (أو "هدف نهائي"، كما يسميه بوستروم)، حتى أن يتم تعريفه لذكاء فائق؟ وفق ما أراه، لا يمكننا أن نثق في رؤية الذكاء الاصطناعي الودود ما لم نتمكن من الإجابة عن هذا السؤال الحاسم.

في أبحاث الذكاء الاصطناعي، عادة ما يكون للأجهزة الذكية هدف واضح ومُحدَّد جيدا، مثلا للفوز بلعبة الشطرنج أو قيادة السيارة إلى الوجهة بشكل قانوني، ينطبق الشيء نفسه على معظم المهام التي نخصصها للبشر، لأن أفق الوقت والسياق معروفان ومحدودان. ولكن الآن نحن نتحدث عن كل مستقبل الحياة في عالمنا، مقيداً فقط بقوانين الفيزياء (لا تزال غير معروفة بالكامل)، لذا فإن تحديد هدف هو مسؤولية مهيبة! وإذا تركنا تأثيرات نظرية الكم جانبا، فإن هدفا مُعرّفا تعريفا جيدا سيحدد كيف يجب أن تترتب جميع الجسيمات في عالمنا عند نهاية الزمن في الفيزياء. إذا ترتبت الجسيمات بتلك الطريقة في وقت أبكر، فإن ذلك الترتيب تقليديا لن يستمر. وأي ترتيب للجسيمات هو المحبّذ على أي حال؟

نحن البشر نميل إلى تفضيل بعض ترتيبات الجسيمات على أخرى؛ مثلاً، نفضل ترتيب مسقط رأسنا كما هو وليس إعادة ترتيب جزيئاته بفعل انفجار قنبلة هيدروجينية. لنفترض أننا نحاول تحديد **وظيفة الخير** ***Goodness function*** التي تربط الرقم بكل ترتيب ممكن للجسيمات في عالمنا، ونحدد كم «خيّرا» هو هذا الترتيب كما نعتقد، ومن ثم نعطي الذكاء الاصطناعي الفائق هدف الوصول بهذه الوظيفة إلى أقصى حدودها. قد يبدو هذا كأنه مقاربة معقولة، فوصف السلوك المُوجَّه نحو الهدف باعتباره تنفيذ الوظائف وصولا إلى حدودها القصوى أمر شائع في مجالات أخرى من العلوم: مثلاً، غالباً ما يضع الاقتصاديون نموذجاً لأفراد وهم يحاولون تحقيق الحدود القصوى لما يسمونه "دالة المنفعة" Utility function، ويُدرِّب العديد من مصممي الذكاء الاصطناعي وكلاءهم الأذكياء لزيادة ما يسمونه "دالة المكافأة" Reward function. عندما نلتزم بالأهداف القصوى لكوزموسنا. ومع ذلك، فإن هذا النهج يشكل كابوساً حوسبياً، لأنه يحتاج إلى تحديد قيمة الخير لكل واحد من أكثر من غوغوبليس من الترتيبات المحتملة للجسيمات الأولية في كوزموسنا- حيث غوغوبليكس يساوي العدد واحد متبوعا بـ 10100 صفر - أي عدد من الأصفار أكبر من عدد الجسيمات في كوننا. فأنّى لنا أن نعرّف دالة الخير هذه لذكاء اصطناعي؟

كما بحثنا آنفا، قد يكون السبب الوحيد الذي يجعلنا نحن البشر ذوي تفضيلات معينة على الإطلاق هو أننا الحل لمشكلة أَمْثَلَة التطور. ومن ثم فإن كل الصفات المعيارية Normative في لغتنا البشرية، مثل "لذيذ" و"عطرة" و "جميل" و"مريح" و"مثير للاهتمام" و"مفيد" و"سعيد" و"جيد"، ترجع أصولها إلى هذه الأمثلة التطورية: لذلك لا يوجد أي ضمان بأن ذكاء اصطناعيا فائق الذكاء سيجدها مُعرَّفة تعريفا صارما. حتى لو تعلم الذكاء الاصطناعي التنبؤ بدقة بتفضيلات بعض البشر، فلن يكون قادرا على حوسبة دالة الخير لمعظم ترتيبات الجسيمات: معظم ترتيبات الجسيمات الممكنة تتوافق مع سيناريوهات كوزموسية غريبة من دون نجوم أو كواكب أو أشخاص بتاتا، وببشر ليس لديهم أي خبرة، فمن الذي يقول إلى أي مدى هي "خيّرة"؟

بالطبع يمكن تعريف بعض الوظائف لترتيب جسيمات الكوزموس تعريفا صارما، بل ونحن على علم بوجود أنظمة مادية تتطور لتحقيق هذه الوظائف إلى أقصى درجة ممكنة. مثلاً، لقد ناقشنا بالفعل عدد الأنظمة التي تتطور لتحقيق أقصى قدر من إنتروبيتها، الأمر الذي في النهاية يؤدي في غياب الجاذبية إلى موت حراري، حيث يكون كل شيء موحداً ولا يتغير لدرجة تبعث على الملل. لذلك، فالإنتروبية ليست بالكاد الشيء الذي نود أن يطلق عليه الذكاء الاصطناعي سمة "خيرة" والسعي إلى أقصى قدر ممكن لها. فيما يلي بعض الأمثلة عن الكميات الأخرى التي يمكن للمرء أن يسعى إلى تحقيق أقصى قدر منها، والتي قد يمكن تعريفها تعريفا صارما من حيث ترتيب الجسيمات:

- جزء من كل مادة في كوننا التي هي في شكل كائن حي معين، مثلا بشرا أو **أيشيريشيا كولاي** *E. coli* (ناجم من تحقيق أقصى قدر من الشمولية التطورية).
- قدرة الذكاء الاصطناعي على التنبؤ بالمستقبل، وهو ما يجادل به الباحث ماركوس هوتر، وهو مقياس جيد للذكاء.
- ما الذي يطلق عليه الباحثون في الذكاء الاصطناعي أليكس ويسنر-غروس Alex Wissner-Gross وكاميرون فرير Cameron Freer مصطلح **إنتروبيا** (وكيل لفرص المستقبل)، والذي يجادلون به هو السمة المميزة للذكاء.
- القدرة الحاسوبية لكوزموسنا.
- التعقيد الخوارزمي لكوزموسنا (عدد البتات اللازمة لوصفه).
- مقدار الوعي في كوزموسنا (انظر الفصل التالي).

ولكن، عندما يبدأ المرء بمنظور الفيزياء، إذ يتكون الكوزموس من جسيمات أولية في الحركة، فمن الصعب أن نرى كيف سيكون تفسير «الخير» واحداً بدلاً من الآخر متميزاً بشكل خاص، لا يزال يتعين علينا تحديد أي هدف من أهداف الكوزموس للوحدة الخاصة بنا والذي يبدو مستحيلا ومرغوبا، إن الأهداف الوحيدة القابلة للبرمجة حالياً والمضمونة لاستمرار احترامه لها حقّاً - نظراً لأن الذكاء الاصطناعي يزداد ذكاء تدريجيّاً- هي الأهداف المعبر عنها بالكميات المادية وحدها، مثل ترتيبات الجسيمات والطاقة والإنتروبية، ولكن، ليس لدينا حالياً أي سبب للاعتقاد أن مثل هذه الأهداف المحتملة ستكون مرغوب فيها لضمان بقاء البشرية.

على النقيض من ذلك، يبدو أننا بشر حادث تاريخي، وليس الحل الأمثل لأي مشكلة فيزيائية جيدة.

هذا يشير إلى أن الذكاء الفائق للذكاء الاصطناعي ذي الهدف المحدد بدقة سيكون قادرا على تحسين تحقيق هدفه من خلال القضاء علينا، وهذا يعني أنه لكي نقرر بحكمة ما يجب القيام به حيال تطوير الذكاء الاصطناعي، فنحن البشر في حاجة ليس فقط لمواجهة التحديات الحوسبية التقليدية، ولكن أيضاً للإجابة عن بعض أكثر الأسئلة فظاعة في الفلسفة، لبرمجة سيارة ذاتية القيادة، نحتاج إلى حل مشكلة العربة التي ستصطدم بها خلال حادث؛ لبرمجة الذكاء الاصطناعي الودي، نحتاج إلى تحديد معنى الحياة، ما هو "معناها"؟ ما "الحياة"؟ ما هي أخلاقيات العمل القصوى؟ بعبارة أخرى: كيف ينبغي أن نسعى جاهدين إلى تشكيل مستقبل الكوزموس الذي نعيش فيه؟ إذا تنازلنا عن السيطرة على ذكاء فائق قبل الإجابة عن هذه الأسئلة بدقة، فمن غير المرجح أن الإجابة التي يتوصل إليها الذكاء الاصطناعي تشملنا نحن البشر، وهذا وقت مناسب لإحياء المناقشات الكلاسيكية حول الفلسفة والأخلاق، بل ويلقي بإلحاح شديد على ضرورة إجراء هذه المناقشة!

الخلاصة:

- أصل السلوك الموجّه نحو الهدف يكمن في قوانين الفيزياء التي تنطوي على التحسين.
- الديناميكا الحرارية لها هدف كامن فيها وهو التبديد Dissipation: زيادة قدر من الفوضى التي تسمى الإنتروبية Entropy.
- الحياة ظاهرة يمكن أن تساعد على التبديد (زيادة الفوضى الكلية) بشكل أسرع من خلال الاحتفاظ بتعقيدها وتكرارها وتزايدها مع زيادة الفوضى في بيئتها.
- يحول التطور الدارويني السلوك الموجّه نحو الهدف من التبديد إلى النسخ.
- الذكاء هو القدرة على تحقيق الأهداف المعقدة.
- نظراً لأننا لا نمتلك دائماً الموارد اللازمة لتنفيذ استراتيجية النسخ المثلى حقاً، فقد طورنا قواعد مفيدة تتمثل في توجيه قراراتنا: مشاعر مثل الجوع والعطش والألم والشهوة والرحمة.
- لذلك لم يعد لدينا هدف بسيط مثل النسخ المتماثل؛ عندما تتعارض مشاعرنا مع هدف جيناتنا، فإننا نطيع مشاعرنا، مثل اللجوء إلى تحديد النسل.
- نحن نبني آلات ذكية بشكل متزايد لمساعدتنا على تحقيق أهدافنا، بقدر ما نبني مثل هذه الآلات لإظهار السلوك الموجّه نحو الأهداف، فإننا نسعى جاهدين لمواءمة أهداف الآلات مع أهدافنا.
- تتضمن مواءمة أهداف الآلات مع أهدافنا ثلاث مشكلات لم يتم حلها بعد: جعل الآلات تتعلمها وتتبناها وتتمسك بها.
- يمكن إنشاء ذكاء اصطناعي يمتلك أي هدف تقريباً، ولكن تقريباً أي هدف طموح قد يؤدي إلى أهداف فرعية من الحفاظ على الذات، والحصول على الموارد، والفضول لفهم العالَم فهما أفضل - قد يؤدي الهدفان الأول والثاني بذكاء اصطناعي فائق إلى التسبب بمشكلات للإنسان، والهدف الأخير قد يمنعه من التمسك بالأهداف التي أعطيناها إياه.
- على الرغم من أن معظم البشر يتفقون على العديد من المبادئ الأخلاقية العامة، فإنه من غير الواضح كيفية تطبيقها على كيانات أخرى، مثل الحيوانات غير البشرية والكيانات المستقبلية المتطورة.
- من غير الواضح الكيفية التي سيطبِّقُ بها الذكاء'الاصطناعي الفائق هدفا أقصى غير مُعرَّف ولا يؤدي إلى القضاء على الإنسانية، لذا فإن هذا وقت مناسب لإعادة إذكاء الأبحاث حول بعض القضايا الشائكة في الفلسفة!

الفصل 8

وعي

لا أستطيع أن أتخيل نظرية متماسكة لكل شيء تتجاهل الوعي.

أندريه ليندي، 2002

يجب أن نسعى جاهدين لتنمية الوعي نفسه - لتوليد أضواء أكثر إشراقا في عالم مظلم بخلاف ذلك.

جوليو تونوني، 2012

رأينا أن الذكاء الاصطناعي قد يساعدنا على خلق مستقبل رائع إذا تمكنا من الإجابة عن بعض من أقدم وأصعب المشكلات في الفلسفة - بحلول الوقت الذي نحتاج فيه إليها. ونواجه الفلسفةَ -مثلما يقول نيك بوستروم- بموعد نهائي. في هذا الفصل، دعونا نستكشف واحداً من أكثر الموضوعات الفلسفية الشائكة على الإطلاق: الوعي.

من يهتم؟

الوعي مثير للجدل، فإذا ذكرت "الوعي" لأحد الباحثين في علم الذكاء الاصطناعي أو اختصاصي علم الأعصاب أو اختصاصي علم النفس، فقد يقلبون أبصارهم تعجبا. وإذا كانوا من معلميك، فقد يشفقون عليك ويحاولون إقناعك بعدم إضاعة وقتك في ما يعتبرونه مشكلة ميئوساً منها وغير علمية. في الواقع، أخبرني صديقي كريستوف كوخ Christof Koch، عالم الأعصاب الشهير الذي يقود معهد ألين لعلوم الدماغ Allen Institute for Brain Science، أنه حُذر ذات مرة من العمل على الوعي قبل أن يحصل على عقده الدائم في

الجامعة - وذلك من قبل فرانسيس كريك نفسه لا غيره الحاصل على جائزة نوبل. إذا بحثت عن كلمة "Consciousness" في قاموس **ماكميلان لعلم النفس** ***Macmillan Dictionary of Psychology*** (عام 1989)، فسيخبرك أنه «لم يكتب عنه ما يستحق القراءة».[1] ولكن كما سأشرح في هذا الفصل فإنني أكثر تفاؤلا!

على الرغم من أن المفكرين تأملوا في سر الوعي لآلاف السنين، فإن ظهور الذكاء الاصطناعي يضيف إلحاحاً مفاجئاً، لا سيما فيما يخص سؤال التنبؤ بالكيانات الذكية التي لديها خبرات موضوعية. فكما رأينا في الفصل الثالث، فإن مسألة ما إذا كان ينبغي منح الآلات الذكية شكلاً من أشكال الحقوق تعتمد بشكل أساسي على ما إذا كانت واعية ويمكن أن تعاني من الألم أو تشعر بالسعادة. وكما ناقشنا في الفصل السابع، يصبح من غير المستحسن صياغة أخلاقيات النفعية القائمة على تضخيم التجارب الإيجابية دون معرفة الكيانات الذكية القادرة التي لديها مثل تلك الأخلاقيات. وكما هو مذكور في الفصل الخامس، قد يفضل البعض أن تكون روبوتاتهم غير واعية لتجنب الشعور بالذنب من امتلاك عبيد. من ناحية أخرى، قد يرغبون في عكس ذلك إذا حمَّلوا عقولهم ليتحرروا من القيود البيولوجية: ففي المحصلة، ما الهدف من تحميل نفسك في روبوت يتحدث ويتصرف مثلك إذا كان مجرد زومبي غير واعٍ، ما أعنيه كونك شخصاً تم تحميله فلا تشعر بالكون كأي شيء؟ ألا يعادل ذلك الانتحار من وجهة نظرك الذاتية، على الرغم من أن أصدقائك ربما لا يدركون أن تجربتك الشخصية قد ماتت؟

بالنسبة إلى مستقبل الحياة الطويلة الأجل في الكوزموس (الفصل السادس)، يصبح فهم ما هو وعي -وما هو ليس بوعي- أمرا محورياً: إذا مكنت التكنولوجيا الحياة الذكية من أن تزدهر عبر كوننا لبلايين السنين، فكيف لنا أن نتأكد من أن هذه الحياة واعية وقادرة على تقدير ما يحدث؟ وإذا لم يكن الأمر كذلك، فهل سيكون ذلك -على حد تعبير الفيزيائي الشهير إروين شرودنغر Erwin Schrödinger- «مسرحية تُمثَّل أمام مقاعد فارغة، وليست هي موجودة لأي شخص، أي بالأصح ليست موجودة»؟[2] بعبارة أخرى، إذا مكّنا أحفادا فائقين تكنولوجيا نعتقد خطأ أنهم واعون، فهل ستكون هذه نهاية العالم الزومبي القصوى، محولا وقفنا الكوزموسي الأكبر إلى لا شيء سوى مضيعة فلكية للفضاء؟

ما هو الوعي؟

تُثير العديد من الحجج حول الوعي نقاشات ساخنة أكثر من أن تسليط الضوء على الموضوع، لأن الخصوم يتحدثون عن أمور مختلفة، أي غير مدركين أنهم يستخدمون تعاريف مختلفة لكلمة الوعي، كما هي الحال مع كلمتي "حياة" و"ذكاء"، لا يوجد تعريف صحيح غير متنازع عليه لكلمة "وعي". بدلاً من ذلك، هناك العديد من التعريفات المتنافسة، بما في ذلك الإحساس Sentience واليقظة Wakefulness والوعي بالذات Self-awareness والوصول إلى المدخلات الحسية والقدرة على دمج المعلومات في رواية.[3] في استكشافنا

لمستقبل الذكاء، نريد أن نأخذ بعين الاعتبار نظرة شاملة وجامعة، لا تقتصر على أنواع الوعي البيولوجي الموجودة حتى الآن. لهذا السبب فإن التعريف الذي قدمته في الفصل الأول، والذي التزمت به طوال هذا الكتاب، هو تعريف واسع جدّاً:

الوعي= **التجربة الذاتية**

بعبارة أخرى، إذا كان الأمر يبدو كما لو كان هو أنت الآن، فأنت واعٍ. إن هذا العريف للوعي بالذات هو الذي يكمن في صميم جميع الأسئلة التي يثيرها الذكاء الاصطناعي في المادة السابقة: هل الكون بروميثيوس أو ألفا غو أو تسلا ذاتية القيادة هو شعور بالكون شيئا ما؟

ولتقدير مدى اتساع نطاق تعريفنا للوعي، لاحظ أنه لا يذكر السلوك أو الإدراك أو الوعي بالذات أو العواطف أو الاهتمام. وبهذا التعريف، فأنت مدرك أيضاً عندما تحلم، على الرغم من أنك تفتقر إلى اليقظة أو الوصول إلى المدخلات الحسية و (من المأمول!) أنك لست تسير وتقوم بأفعال وأنت نائم. بالمثل، أي نظام يشعر بالألم هو نظام واعٍ بهذا المعنى، حتى إذا لم يكن بإمكانه أن يتحرك. إن تعريفنا يترك الباب مفتوحاً أمام احتمال بأن تكون بعض أنظمة الذكاء الاصطناعي المستقبلية واعية أيضاً، حتى لو كانت موجودة فقط كبرمجيات ولا ترتبط بأجهزة استشعار أو أجسام روبوتية.

وبمثل هذا التعريف، من الصعب عدم الاهتمام بالوعي. فكما يقول يوفال نوح هراري في كتابه هومو ديوس Homo Deus:[4] «إذا أراد أي عالِم أن يجادل في أن التجارب الشخصية الذاتية غير ذات صلة، فإن التحدي الذي يواجهه هو شرح لماذا يكون التعذيب أو الاغتصاب خطأً من دون الرجوع إلى أي تجربة ذاتية». و من دون مثل هذا المرجعية، فإن الأمر جميعه لا يعدو أن يكون مجرد مجموعة من الجسيمات الأولية تتحرك وفقاً لقوانين الفيزياء - فما الخطأ في ذلك؟

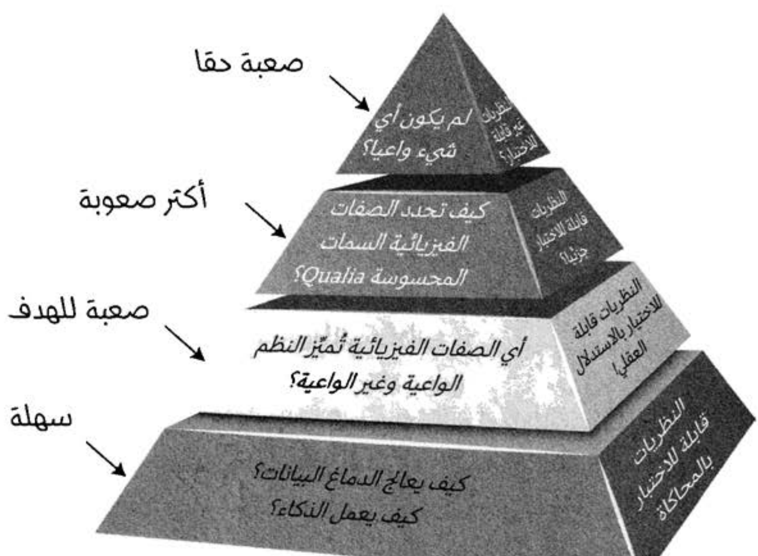
ما المشكلة؟

إذاً هو بالتحديد الذي لا نفهمه بخصوص الوعي؟ قلة من أعملوا الفكر حول هذا السؤال أكثر من ديفيد تشالمرز David Chalmers، الفيلسوف الأسترالي الشهير الذي نادراً ما شوهد من دون ابتسامة مرحة وسترة جلدية سوداء - أحبتها زوجتي كثيراً لدرجة أنها أهدتني سترة مشابهة في عيد الميلاد. لقد تبع قلبه إلى دراسة الفلسفة على الرغم من أنه حقق المراتب الأولى في الأولمبياد الدولي للرياضيات - وعلى الرغم من حقيقة أن درجة B الوحيدة التي حصل عليها في الكلية، محطمة سجله المؤلف تماما من الدرجات A فيما عدا ذلك. في الواقع، من الواضح أنه لم يحبط بالمعوقات أو الجدال، ودائماً يدهشني من قدرته على الاستماع بأدب إلى النقد المُضلّل لعمله وعدم شعوره بالحاجة إلى الرد على ذلك.

كما أكد ديفيد، هناك حقاً لغزان منفصلان للعقل. أولاً، هناك سر كيف يعالج الدماغ المعلومات، التي يطلق عليها ديفيد المشكلات "السهلة". فمثلا، كيف يعالج المخ ويفسر ويستجيب للمدخلات الحسية؟ وكيف يمكنه الإبلاغ عن حالته الداخلية باستخدام اللغة؟ على الرغم من أن هذه الأسئلة هي في الواقع بالغة الصعوبة، غير أنها من حيث التعريف الذي نستخدمه ليست أسرارا من أسرار الوعي، ولكنها أسرار ذكاء: فهي تسأل كيف يتذكر الدماغ ويحسب ويتعلم. إضافة إلى ذلك، رأينا في الجزء الأول من الكتاب كيف بدأ باحثو الذكاء الاصطناعي في إحراز تقدم كبير في حل العديد من هذه "المشكلات السهلة" مع الآلات - بدءاً من اللعبة غو Go إلى قيادة السيارات، وتحليل الصور ومعالجة اللغة الطبيعية.

ثم هناك سر آخر يدور حول لماذا لديك تجربة ذاتية، التي يطلق عليها ديفيد المشكلة الصعبة. فعندما تقود سيارة، فإنك تشعر بتجربة بالألوان والأصوات والعواطف والشعور بالذات. ولكن لماذا تشعر بتجربة أي شيء على الإطلاق؟ هل تشعر السيارة ذاتية القيادة بتجربة أي شيء على الإطلاق؟ إذا كنت تتسابق ضد سيارة ذاتية القيادة، فأنت تقوم بإدخال المعلومات من أجهزة الاستشعار ومعالجتها وإخراج أوامر للمحرك. لكن تجربة القيادة تجربة ذاتية هي شيء منفصل منطقياً - وهو أمر اختياري، وإذا كان الأمر كذلك، فما الذي يسببه؟

أتعامل -أنا- مع مشكلة الوعي الصعبة هذه من وجهة نظر الفيزياء. فمن منظوري، الشخص الواعي هو مجرد طعام معاد ترتيبه، فلماذا يكون ترتيبٌ ما واعيا، ولكن ليس الآخر؟



الشكل 1.8: فهم العقل ينطوي على سلسلة هرمية من المشكلات. ما يدعوه ديفيد تشالمرز مشكلات "سهلة" يمكن طرحه من دون ذكر التجربة الذاتية. إن الحقيقة الواضحة المتمثلة في أن بعض -وليس كل- النظم الفيزيائية واعية تطرح ثلاثة أسئلة مستقلة. إذا كانت لدينا نظرية للإجابة عن سؤال يُحدّد "مشكلة صعبة للهدف"، فيمكن عندها اختبارها تجريبياً. إذا نجح الأمر، فيمكننا البناء عليه لمعالجة الأسئلة الأكثر صرامة المذكورة آنفاً.

إضافة إلى ذلك، تُخبرنا الفيزياء أن الطعام هو ببساطة عدد كبير من الكواركات والإلكترونات، مرتبة بطريقة معينة. إذاً، أي ترتيبات الجسيمات واعية وأيها غير واعٍ؟

ما يعجبني في منظور الفيزياء هذا هو أنه يحول المشكلة الصعبة التي واجهناها كبشر منذ آلاف السنين إلى إصدار أكثر تركيزاً وهو أسهل في التعامل معه بالمنهج العلمي. فبدلاً من البدء بمشكلة صعبة تتمثل في سبب شعور ترتيب الجسيمات بالوعي، دعنا نبدأ بحقيقة صعبة مفادها أن بعض ترتيبات الجسيمات تشعر بالوعي في حين أن البعض الآخر لا يشعر بذلك، مثلا. أنت تعلم أن الجسيمات التي تشكل دماغك هي في ترتيب واعٍ الآن، ولكن ليس عندما تكون في نوم عميق بلا أحلام.

يؤدي منظور الفيزياء هذا إلى ثلاثة أسئلة صعبة مستقلة حول الوعي، كما هو مبين في الشكل 8.1. أولا، أي خصائص ترتيب الجسيمات هي التي تحدث الفرق؟ على وجه التحديد، ما هي الخصائص الفيزيائية التي تميز الأنظمة الواعية وغير الواعية؟ إذا استطعنا الإجابة عن ذلك، فيمكننا معرفة أي من أنظمة الذكاء الاصطناعي واعية. وفي المستقبل القريب، يمكن أن تساعد الإجابةُ الأطباءَ في غرفة الطوارئ على تحديد أي المرضى الذين لا يستجيبون هم واعون بالفعل.

ثانياً، كيف تحدد الخواصُ الفيزيائية ماهية شكل التجربة؟ على وجه التحديد، ما الذي يحدد السمات المحسوسة Qualia، أي لبنات بناء الوعي، مثلا: احمرار الوردة، صوت الصنج، رائحة شريحة لحم، أو طعم اليوسفي أو شكة الإبرة؟*

ثالثاً، لماذا هناك ما هو واعٍ؟ بعبارة أخرى، هل هناك بعض التفسيرات العميقة غير المكتشفة حول سبب كون كميات من المادة واعية، أم أنها مجرد حقيقة وحشية لا يمكن تفسيرها حول الطريقة التي يعمل بها العالم؟

وصف عالم الحاسوب سكوت آرونسون -وهو زميل سابق لي من المعهد MIT- بمرح السؤالَ الأول "المشكلة الصعبة للهدف" Pretty hard problem (اختصارا: المشكلة PHP)، كما فعل ديفيد تشالمرز. وبهذه الروح، دعنا نسمِّ المشكلتين الأخريين: "المشكلة الأكثر صعوبة" Even harder problem (اختصارا: المشكلة EHP) و"المشكلة الصعبة حقاً" Really hard problem (اختصارا: المشكلة RHP)، كما هو موضح في الشكل 8.1.**

* أستخدم "**السمات المحسوسة**" ***Qualia*** وفقاً لتعريف القاموس لتعني الحالات الفردية من التجربة الذاتية - أي إنها تعني التجربة الشخصية نفسها، وليس أي مادة مقترحة تُسبِّب التجربة. انتبه، فالبعض يستخدمون الكلمة بشكل مختلف.

** وصفت "المشكلة الصعبة حقاً" Really hard problem في الأصل بـ "المشكلة الصعبة للهدف" Pretty hard problem (اختصارا: المشكلة PHP). لكن بعد أن عرضت هذا الفصل على ديفيد تشالمرز، أرسل إليَّ عبر البريد الإلكتروني اقتراحاً ذكياً باستبدالها بـ "**المشكلة الصعبة حقاً**" ***Really hard problem***، لمطابقة ما يعنيه حقاً، إذ كتب: «لما كانت المشكلتان الأولى والثانية (على الأقل كما هو موضح في هذا الفصل) ليستا جزءاً من المشكلة الصعبة كما تصورتُها في حين أن المشكلة الثالثة هي جزء منها، ربما يمكنك استخدام 'صعب حقاً' Really hard بدلاً من 'صعب جدّاً' Very hard للمشكلة الثالثة لتطابق استخدامي.

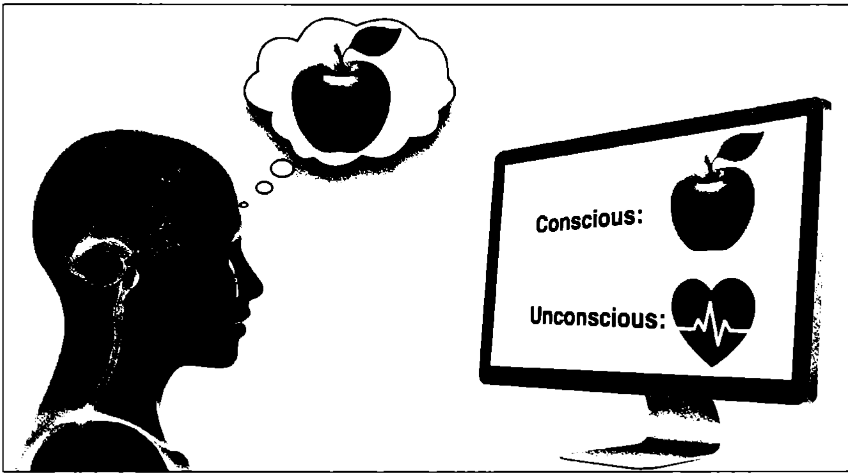
هل الوعي وراء حدود العلم؟

عندما يخبرني البعض أن أبحاث الوعي هي مضيعة للوقت بشكل ميؤوس منه، فإن الحجة الرئيسية التي يقدمونها هي أنها "غير عِلْمية" Unscientific وستظل كذلك دائماً. ولكن هل هذا صحيح حقا؟ روّج الفيلسوف النمساوي البريطاني كارل بوبر Karl Popper القول المأثور المقبول الآن على نطاق واسع: «إذا لم يكن قابلا للدحض Falsifiable، فهو ليس علماً». وبعبارة أخرى، فإن العلم يدور حول اختبار النظريات مقارنة بالملاحظات المرصودة: إذا كانت النظرية لا يمكن اختبارها حتى من حيث المبدأ، إذاً من المستحيل منطقيا دحضها؛ مما يعني حسب تعريف بوبر أنه غير علمي.

إذاً هل هناك نظرية علمية تجيب عن أي من أسئلة الوعي الثلاثة في الشكل 1.8؟ رجاء اسمح لي أن أحاول إقناعك بأن الإجابة هي: نعمٌ مدوية!، على الأقل بالنسبة إلى المشكلة الصعبة للهدف: «ما هي الخصائص الفيزيائية التي تميز الأنظمة الواعية وغير الواعية؟» لنفترض أن شخصاً ما لديه نظرية- آخذين بعين الاعتبار أي نظامٍ مادي- تجيب عن سؤال حول ما إذا كان النظام مدركاً بـ "نعم" أو "لا" أو "غير متأكد". دعنا نربط عقلك بجهاز يقيس بعضا من معالجة المعلومات في أجزاء مختلفة من دماغك، ودعونا نرسل هذه المعلومات إلى برنامج حاسوبي يستخدم نظرية الوعي للتنبؤ بأي جزء من تلك المعلومات هو واعٍ، ويقدم لك تنبؤاته في الوقت الحقيقي على الشاشة، كما هي الحال في الشكل 2.8. أولاً تفكر في تفاحة. تبلغك الشاشة بوجود معلومات حول تفاحة في دماغك وأنت واعٍ بتلك المعلومات، ولكن هناك أيضاً معلومات في جذع الدماغ Barinstem عن نبضك الذي أنت غير مدرك له. هل يُدهشك هذا؟ على الرغم من أن أول تنبؤين للنظرية صحيحان، فإنك قررت القيام ببعض الاختبارات الأكثر صرامة. تفكر في أمك فيخبرك الحاسوب أن هناك معلومات في دماغك عن والدتك ولكنك غير مدرك لذلك. تنبأت النظرية تنبؤا غير صحيح؛ مما يعني أنها تُستبعد وتذهب في مزبلة تأريخ العلم جنبا إلى جنب مع الميكانيكا الأرسطية Aristotelian mechanics، والأثير المضيء Luminiferous aether، ومركزية الأرض، والعديد من الأفكار الفاشلة الأخرى. النقطة الأساسية: على الرغم من أن النظرية كانت خاطئة، فإنها كانت علمية! **لو** لم تكن علمية، لما تمكنت من اختبارها واستبعادها.

قد ينتقد البعض: ما هذا الاستنتاج، ويقول إنه لا يوجد لديه دليل على ما أنت واعٍ به، أو حتى إذا ما كنت واعٍ على الإطلاق: على الرغم من أنهم سمعوك تقول إنك على وعي، غير أنه من الممكن تخيل أن زومبيا لاوعيا قد يقول الشيء نفسه. لكن هذا لا يجعل نظرية الوعي هذه غير علمية، لأنها قد تتبادل الأماكن معك وتختبر ما إذا كانت تتنبأ بتجاربها الواعية بشكل صحيح.

من ناحية أخرى، إذا رفضت النظرية تقديم أي تنبؤات، وكان الجواب هو: «غير متأكد» عن كل استفسار، من ثم تكون غير قابلة للاختبار ومن ثم غير علمية. قد يحدث هذا لأنها قابلة للتطبيق فقط في بعض الحالات، أو لأن الحسابات المطلوبة يصعب تنفيذها عمليا.



الشكل 2.8: لنفترض أن حاسوبا يقيس المعلومات التي يعالجها دماغك ويتوقع أي جزء منها تُدركه وفقاً لنظرية الوعي. يمكنك اختبار هذه النظرية عِلمياً عن طريق التحقق من صحة تنبؤاتها ومطابقتها لتجربتك الذاتية.

أو لأن مجسات الدماغ ليست جيدة. تميل نظريات اليوم العلمية الأكثر شيوعاً إلى أن تكون في مكان ما في الوسط؛ مما يعطي إجابات قابلة للاختبار لبعض أسئلتنا وليس كلها. مثلاً، سترفض نظريتنا الأساسية في الفيزياء الإجابة عن أسئلة حول الأنظمة التي تكون في الوقت نفسه صغيرة جدّاً (تتطلب ميكانيكا الكم) وثقيلة جدّاً (تتطلب النسبية العامة)، لأننا لم نتوصل بعد إلى أي المعادلات الرياضياتية يجب أن نستخدمها في هذه الحالة. سوف ترفض هذه النظرية الأساسية أيضاً التنبؤ بالكتل الدقيقة لجميع الذرات الممكنة - في هذه الحالة، نعتقد أن لدينا المعادلات الضرورية. لكننا لم نتمكن من حساب حلولها بدقة، وكلما كانت النظرية مُخاطِرة -بتعريض نفسها للتمحيص- ووضع تنبؤات قابلة للاختبار، كانت أكثر فائدة، وأخذناها على محمل الجد أكثر إذا نجت من جميع محاولاتنا لقتلها. نعم، يمكننا فقط اختبار بعض تنبؤات نظريات الوعي، ولكن هذه هي الحال في جميع النظريات الفيزيائية. لذلك دعونا لا نضيع الوقت في الشكوى مما لا يمكننا اختباره، ونركز على ما يمكننا اختباره!

باختصار، إن أي نظرية تتنبأ بأي من الأنظمة الفيزيائية هي أنظمة واعية (المشكلة الصعبة للهدف) هي نظرية علمية، مادامت تستطيع التنبؤ بأي من عمليات الدماغ هي عمليات واعية. ولكن، تصبح قضية القابلية للاختبار أقل وضوحاً بالنسبة للأسئلة العليا في الشكل 1.8. ما الذي يعنيه أن تتنبأ النظرية بتجربة اللون الأحمر تجربة ذاتية؟ وإذا زعمت النظرية أنها تشرح سبب وجود شيء من هذا القبيل في الوعي في المقام الأول، فكيف يمكنك اختبارها تجريبياً؟ فقط لأن هذه الأسئلة صعبة لا يعني أننا يجب أن نتجنبها، وسنعود إليها بالفعل. لكن عند مواجهة العديد من الأسئلة التي لم تتم الإجابة عنها، أعتقد

أنه من الحكمة معالجة السؤال الأسهل أولاً. لهذا السبب، فإن بحثي في الوعي في المعهد MIT يركز بشكل مباشر على قاعدة الهرم في الشكل 8.1. لقد ناقشت هذه الاستراتيجية مؤخراً مع زميلي في الفيزياء بييت هات Piet Hut من جامعة برينستون Princeton University، الذي قال مازحا إن محاولة بناء الجزء العلوي من الهرم قبل القاعدة ستكون مثل القلق بشأن تفسير ميكانيكا الكم قبل اكتشاف معادلة شرودنغر، الأساس الرياضي الذي يتيح لنا التنبؤ بنتائج تجاربنا.

عند مناقشة ما هو خارج نطاق العلم، من المهم أن تتذكر أن الإجابة تعتمد على الوقت! قبل أربعة قرون، أُعجب غاليليو غاليلي بنظريات الفيزياء القائمة على الرياضيات لدرجة أنه وصف الطبيعة بأنها «كتاب مكتوب بلغة الرياضيات». إذا ألقى حبة عنب وبندقة، كان بإمكانه أن يتنبأ بدقة بأشكال مساراتهما ومتى سترتطمان بالأرض. ومع ذلك، لم يكن لديه أدنى فكرة عن سبب كون إحداهما خضراء والأخرى بنية، أو سبب كون إحداهما طرية والأخرى صلبة - كانت هذه الجوانب من العالَم بعيدة عن متناول العلم في ذلك الوقت. لكن ليس إلى الأبد! فعندما اكتشف جيمس كليرك ماكسويل James Clerk Maxwell معادلاته المُسمَّاة باسمه في عام 1861، صار من الواضح أن الضوء والألوان يمكن فهمهما أيضاً رياضياتيا. نحن نعلم الآن أن معادلة شرودنغر المذكورة آنفا، المُكتشفة في عام 1925، يمكن استخدامها للتنبؤ بجميع خصائص المادة، بما في ذلك ما إذا كانت طرية أو صلبة. وبينما مكّن التقدم النظري -أكثر من أي وقت مضى- التنبؤات العلمية، فقد مكّن التقدم التكنولوجي من إجراء المزيد من الاختبارات التجريبية: فكل شيء تقريباً نستخدمه الآن لدراسة الظواهر من التلسكوبات أو المجاهر أو مصادمات الجسيمات كان في يوم ما يقع خارج نطاق العلم. وبعبارة أخرى، اتسع نطاق العلوم بشكل كبير منذ أيام غاليليو، من جزء ضئيل من جميع الظواهر إلى نسبة مئوية كبيرة منها، بما في ذلك الجسيمات تحت الذرية، والثقوب السوداء وأصول الكوزموس منذ 13.8 بليون سنة. وهذا يطرح السؤال: ماذا تبقى؟

بالنسبة إلي، الوعي هو الفيل في الغرفة. فأنت لا تعرف فقط أنك واعٍ، لكن هو ما تعرفه على وجه اليقين التام - كل شيء آخر هو استدلال، كما أشار رينيه ديكارت René Descartes ذات يوم في زمن غاليليو. هل سيجلب التقدم النظري والتكنولوجي الوعي في نهاية المطاف إلى مجال العلوم؟ لا نعرف، تماماً كما لم يعرف غاليليو ما إذا كنا سنفهم يوماً ما الضوء والمادة.* الشيء الوحيد المضمون: لن ننجح إذا لم نحاول! هذا هو السبب في أنني والعلماء الآخرين في جميع أنحاء العالم نحاول جاهدين صياغة واختبار نظريات الوعي.

* إذا كان واقعنا المادي رياضياتيا تماماً (قائماً على المعلومات، بعبارة فضفاضة)، كما بحثت في كتابي **كوننا الرياضياتي** ***Our Mathematical Universe***، فلا يقع أي جانب من الواقع - ولا حتى من الوعي - خارج نطاق العلوم. في الواقع، إن مشكلة الوعي الصعبة حقاً هي -من هذا المنظور- مثل مشكلة كيف يمكن نشعر بشيء رياضياتي كما لو كان ماديا: إذا كان جزء من بنية الرياضيات واعياً، فستستشعر ببقية الأجزاء باعتبارها عالمها المادي الخارجي.

أدلة تجريبية حول الوعي

يجرى الكثير من معالجة المعلومات في رؤوسنا الآن. أي منها واعٍ وأي منها ليس كذلك؟ قبل استكشاف نظريات الواعي وما تتنبأ به، دعونا نلقِ نظرة على ما علمتنا إياه التجارب حتى الآن، بدءاً من الملاحظات الرصدية لقياسات الدماغ باستخدام الطرق منخفضة التقنية وصولا إلى التقنيات الحديثة.

أي السلوك واعٍ؟

إذا ضربت 32 في 17 في رأسك، فأنت واعٍ بالعديد من الإجراءات الداخلية لحساباتك. لكن لنفترض أني أريك بدلاً من ذلك صورة لألبرت آينشتاين وأطلب إليك أن تخبرني باسم موضوعه. كما رأينا في الفصل الثاني. هذه أيضاً مهمة حسابية: يقوم عقلك بتقييم وظيفة يكون مدخلاتها معلومات من عينيك عن عدد كبير من ألوان البكسلات ناتجها عبارة عن معلومات للعضلات التي تتحكم في الفم والأسلاك الصوتية. يطلق علماء الحاسوب على هذه المهمة "تصنيف الصور" Image classification متبوعة "بتخليق (إنشاء) الكلام" Speech synthesis. على الرغم من أن هذا الحوسبة أكثر تعقيدا من مهمة الضرب، يمكنك القيام بذلك بشكل أسرع بكثير، وظاهريا من دون جهد، ودون أن تكون واعيا بتفاصيل الكيفية التي يمكنك بها أن تفعل ذلك. تتألف تجربتك الذاتية بمجرد النظر إلى الصورة، وتجربة شعور بالتعرف وسماع نفسك تقول: «آينشتاين».

عرف علماء النفس منذ فترة طويلة أنه بإمكانك فهم مجموعة واسعة من المهام والسلوكيات الأخرى دون أن تكون واعيا أيضاً، بدءاً من رد فعل رمشة العين وحتى التنفس ومد الذراع والقبض باليد والحفاظ على توازنك. عادةً ما تكون على دراية واعية بما قمت به، وليس كيفية قيامك بذلك. من ناحية أخرى، فإن السلوكيات التي تنطوي على أوضاع غير مألوفة، أو ضبط النفس، أو القواعد المنطقية Logical rule المعقدة، أو التفكير المنطقي أو التركيبات اللغوية تميل إلى أن تكون واعية. وهي تعرف بالارتباط السلوكي للوعي Behavioral correlates of consciousness، و ترتبط ارتباطاً وثيقاً بطريقة التفكير التي تتسم ببذل الجهد والبطيئة والمسيطر عليها والتي يطلق عليها علماء النفس System 2 "النظام 2". [5]

من المعروف أيضاً أنه يمكنك تحويل العديد من الإجراءات الروتينية من الوعي إلى اللاوعي عبر الممارسة المُكثّفة. مثلا: المشي والسباحة وركوب الدراجات والقيادة والطباعة والحلاقة وربط الأحذية وألعاب الحاسوب وعزف البيانو. [6] بالطبع، من المعروف جيدا أن خبراء يؤدون مهام تخصصاتهم بشكل أفضل عندما يكونون في حالة من "التدفق" Flow مدركين فقط لما يحدث عند المستوى الأعلى وبلا وعي بتفاصيل المستوى المنخفض لكيفية القيام بذلك. مثلاً، حاول قراءة الجملة التالية مع إدراكك

الواعي لكل حرف، كما كنت تقرأ عندما كنت تتعلم القراءة، هل يمكنك أن تشعر بمدى البطء مقارنةً بالوقت الذي تكون فيه واعيا للنص على مستوى الكلمات أو الأفكار؟

في الواقع، لا يبدو أن معالجة المعلومات اللاواعية ممكنة فحسب، بل إنها أيضاً القاعدة أكثر منها الاستثناء. تشير الأدلة إلى أنه من بين نحو 107 بتات من المعلومات التي تدخل دماغَنا كل ثانية من أجهزتنا الحسية، يمكننا أن ندرك فقط جزءا ضئيلا منها، بتقديرات تتراوح من 10 إلى 50 بت.[7] هذا يشير إلى أن معالجة المعلومات التي ندركها بوعي هي مجرد غيض من فيض.

إن هذه القرائن، مجتمعة، دفعت ببعض الباحثين إلى اعتبار أن المعالجة الواعية للمعلومات هي المدير التنفيذي لأذهاننا، إذ تتناول فقط أهم القرارات التي تتطلب تحليلاً معقداً للبيانات من جميع أنحاء الدماغ.[8] وهذا يفسر السبب، تماماً مثل رئيس تنفيذي لشركة، الدماغ لا يريد عادةً أن يتشتت انتباهه بمعرفة كل شيء يقوم به مرؤوسوه. لكن يمكنه أن يعرفها إذا رغب في ذلك. ولرؤية هذا الاهتمام الانتقائي أثناء تطبيقه، أعد النظر في تلك الكلمة "رغب"مرة أخرى: ثبت بصرك على نقطة فوق "غ"، ومن دون تحريك عينيك، حول انتباهك من النقطة إلى حرف كله ومن ثم إلى الكلمة كلها. على الرغم من أن المعلومات من شبكية عينك بقيت كما هي، فإن تجربتك الواعية تغيرت، واستعارة الرئيس التنفيذي تشرح أيضاً لماذا تصبح الخبرة لاواعية: بعد جهد تعلم القراءة والكتابة، يفوض المدير التنفيذي هذه المهام الروتينية لمفوَّضين لاواعين ليركِّز على تحديات المستوى الأعلى.

أين الوعي؟

أشارت تجارب وتحليلات حاذقة إلى أن الوعي لا يقتصر فقط على سلوكيات معينة، ولكن أيضاً على أجزاء معينة من الدماغ، من هم المشتبه فيهم الرئيسيون؟ جاء الكثير من القرائن الأولى من مرضى يعانون إصابات في الدماغ: تلفاً موضعياً في الدماغ ناتجاً عن الحوادث أو السكتة الدماغية أو الأورام أو الالتهابات. ولكن في كثير من الأحيان لم تكن هذه الأدلة حاسمة. مثلا: حقيقة أن إصابات الجزء الخلفي من الدماغ قد تسبب العمى هل تعني أن هذا هو موقع الوعي البصري، أم أن المعلومات المرئية تمر عبر هذا الطريق إلى المكان الذي ستغدو فيه فيما بعد واعية، تماماً كما تمر عبر العين؟

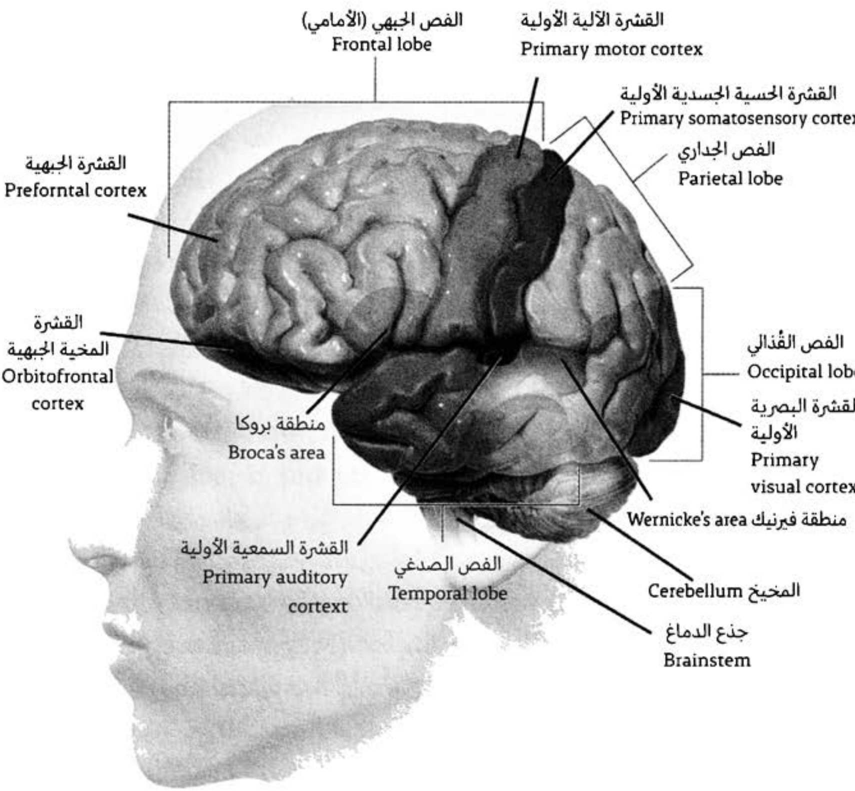
على الرغم من أن الإصابات والتدخلات الطبية لم تحدد مواقع التجارب الواعية، فإنها ساعدت على تضييق نطاق الخيارات. مثلاً، على الرغم من أنني أعاني ألما في يدي أعلم أنه يحدث هناك بالفعل، فإن تجربة الألم يجب أن تحدث في مكان آخر، لأن الجراح أوقف ألم يدي دون فعل أي شيء ليدي: بمجرد تخدير الأعصاب في كتفي. إضافة إلى ذلك، فإن بعض مبتوري الأطراف يعانون الألم الوهمي Phantom pain الذي يبدو وكأنه ناجم من يدهم غير الموجودة. وكمثال آخر، لاحظت ذات مرة أنني عندما أنظر

فقط بعيني اليمنى فإن أفقد جزءا من مجال نظري - حدّد الطبيب أن شبكية العين كانت قد انفصلت عن مكانها فأعاد تثبيتها. وعلى النقيض من ذلك، فإن المرضى المصابين بجروح دماغية معينة يعانون **الإهمال الحيزي النص في *Hemineglect***، إذ يفتقدون أيضاً المعلومات من نصف مجالهم البصري. لكنهم لا يدركون حتى أنها مفقودة - مثلاً، عدم ملاحظة الطعام وتناوله في النصف الأيسر من صحنهم. يبدو الأمر كما لو أن الوعي بنصف عالمهم قد اختفى، ولكن هل من المفترض أن تولد مناطق الدماغ التالفة هذه التجربة المكانية، أم أنها تُغذي مواقع الوعي بالمعلومات المكانية، مثلما فعلت شبكية العين؟

في عام 1930 وجد جراح الأعصاب الأمريكي-الكندي الرائد ويلدر بن Wilder Pen أن مرضى جناح جراحة الأعصاب أبلغوا عن لمس أجزاء مختلفة من الجسم عندما حفَّز مناطق محددة من الدماغ في ما يسمى الآن **بالقشرة الحسية الجسدية *Somatosensory cortex*** (الشكل 3.8).[9] ولاحظ أيضاً أنهم حركوا لاإراديا أجزاء مختلفة من الجسم عندما حفز مناطق الدماغ في ما يسمى الآن القشرة الحركية Motor cortex. ولكن هل يعني ذلك أن معالجة المعلومات في مناطق المخ هذه تتوافق مع وعي اللمس والحركة؟

لحسن الحظ، فإن التكنولوجيا الحديثة توفر لنا الآن الكثير من الأدلة التفصيلية. على الرغم من أننا لا نزال بعيدين عن قياس كل إشارة عصبية تُطلق من جميع ما يقرب من مئة بليون خلية عصبية، فإن تكنولوجيا قراءة المخ تتقدم بسرعة، إذ تتضمن تقنيات ذات أسماء مخيفة مثل التصوير fMRI والتخطيط EEG والتخطيط MEG والتخطيط ECoG واختبار ePhys واستشعار جهد الفلوريسنت (الفلورة) Fluorescent voltage sensing. يقيس التصوير بالرنين المغناطيسي الوظيفي fMRI (اختصارا لـ Functional magnetic resonance imaging) الخواص المغناطيسية لنواة الهيدروجين لرسم خريطة ثلاثية الأبعاد لعقلك لكل ثانية تقريباً، بدقة تصل إلى مليمتر واحد. أما التخطيط EEG (تخطيط كهربية الدماغ Electroencephalography) والتخطيط MEG (تخطيط مغناطيسية الدماغ Magnetoencephalography) فيقيسان المجال الكهربائي والمغناطيسي خارج رأسك لرسم خريطة لعقلك عدة آلاف من المرات في الثانية، ولكن بدقة منخفضة، غير قادرة على التمييز بين السمات الأصغر من بضعة سنتيمترات. إذا كنت من النوع القلق، فستحبذ هذه التقنيات الثلاث لأنها غير باضعة Noninvasive. إذا كنت لا تمانع في فتح جمجمتك، فلديك خيارات إضافية، يتضمن التخطيط ECoG (تخطيط كهربية قشر الدماغ Electrocorticography) وضع مئات الأسلاك على سطح دماغك، في حين أن الاختبار ePhys (الفيزيولوجيا الكهربية Electrophysiology) يتضمن إدخال أسلاك دقيقة Microwires -تكون في بعض الأحيان أرق من شعر الإنسان- في عمق الدماغ لتسجيل الفولتية من أكثر من ألف موقع في وقت واحد. يقضي العديد من

مرضى الصرع أياماً في المستشفى فيما يستخدم الأطباء التخطيط ECoG للتعرف على أي جزء من الدماغ يؤدي إلى نوبات ومن ثم يجب استئصاله، ويسمحون لعلماء الأعصاب بإجراء تجارب على الوعي في أثناء ذلك. أخيراً، يشمل اختبار استشعار جهد الفلورسنت الناجم عن الخلايا العصبية تعديلها جينيا لينبعث منها ومضات ضوء عند إطلاق الإشارة العصبية (القدح) Firing؛ مما يسمح بقياس نشاطها تحت المجهر. من بين جميع التقنيات، لاستشعار جهد الفلورسنت القدرة على مراقبة أكبر عدد من الخلايا العصبية بسرعة، على الأقل في الحيوانات ذات العقول الشفافة - مثل الدودة (الربداء الرشيقة) سي. إيليجانس C. elegans بعصبوناتها الـ 302. ويرقة سمكة الزرد Zebrafish بعصبوناتها التي تعادل 100,000 تقريباً.



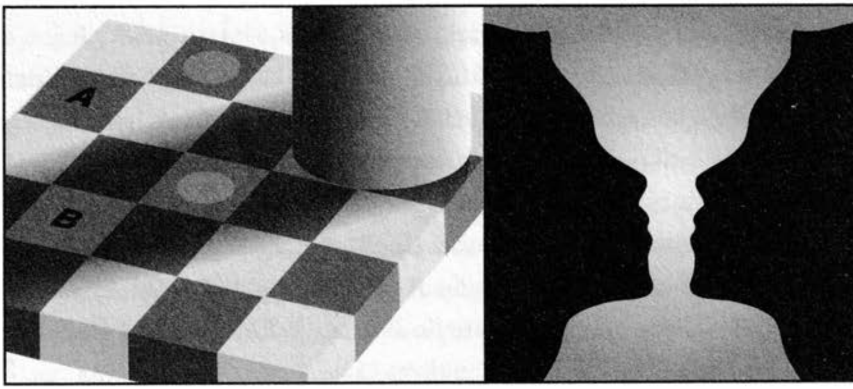
الشكل 3.8: تشارك القشرة البصرية والسمعية والحسية الجسدية والحركية في الرؤية والسمع والشعور باللمس وتفعيل الحركة، على التوالي - لكن هذا لا يثبت أنها المنطقة التي يتولد فيها **الوعي** بالرؤية والسمع واللمس والحركة. في الواقع، تشير الأبحاث الحديثة إلى أن القشرة البصرية الأولية Primary visual cortex غير واعية نهائيا، وكذلك هو المخيخ وجذع المخ. إهداء من لاشينا (www.lachina.com).

وعلى الرغم من أن فرانسيس كريك Francis Crick حذر كريستوف كوتش Christof Koch من دراسة الوعي، فإن كريستوف رفض الاستسلام وفاز فرانسيس في النهاية. في عام 1990، كتبا ورقة بحثية حول ما أطلقا عليه "الارتباطات العصبية للوعي" Neural correlates of consciousness (اختصارا: الارتباطات NCCs)، متسائلين عن أيٍّ من عمليات الدماغ هي التي تتوافق مع التجارب الواعية. لآلاف السنين، كان المفكرون يَصِلون إلى المعلومات التي تدور في أدمغتهم فقط من خلال تجربتهم وسلوكهم الذاتيين. وأشار كريك وكوتش إلى أن تقنيات قراءة المخ صارت توفر فجأة وصولاً مستقلاً إلى هذه المعلومات؛ مما يسمح بدراسة علمية لمعرفة أيٍّ من معالجات المعلومات تتوافق مع التجربة الواعية. من المؤكد أن القياسات المُستندة إلى التكنولوجيا حولت الآن البحث عن الارتباطات NCCs إلى جزء رئيسي من علم الأعصاب، الجزء الذي نشرت آلاف من أوراقه البحثية في أفضل الدوريات المرموقة.[10]

ما هي الاستنتاجات حتى الآن؟ لإلقاء نظرة على عمل مُختبر الارتباطات NCC، دعنا نتساءل أولاً عما إذا كانت شبكية عينك واعية أم أنها مجرد نظام زومبي يسجل المعلومات المرئية ويعالجها ويرسلها إلى نظام ما في دماغك حيث تتولد تجربتك البصرية الذاتية. إلى اليسار في الشكل 4.8. أي مربع أغمق: المربع A أو B؟ لا، إنها في الواقع ألوان متماثلة، ويمكنك التحقق من ذلك بالنظر إليهما من ثقب صغير بين أصابعك. هذا يثبت أن تجربتك المرئية لا يمكن أن توجد بالكامل في شبكية عينك، لأنها إذا كانت كذلك، فسيبدو المربعان متماثلين.

انظر الآن إلى الجانب الأيمن من الشكل 4.8. هل ترى امرأتين أم مزهرية؟ إذا دققت النظر لوقت طويل بما فيه الكفاية، فستشاهد كل منهما ذاتياً على التناوب، على الرغم من أن المعلومات التي تصل إلى شبكية العين تظل كما هي. بقياس ما يحدث في دماغك خلال الحالتين، يمكن للمرء أن يستخلص ما الذي يسبب ذلك الفرق- وبالتأكيد ليس ذلك بالشبكية التي تتصرف بشكل متماثل في كلتا الحالتين.

تأتي الضربة القاضية لفرضية الشبكية الواعية من تقنية تسمى "التثبيط المستمر للوميض" Continuous flash suppression التي ابتكرها كريستوف كوخ، وستانيسلاس ديهان Stanislas Dehaene، والمتعاونون معها: اكتشفنا أنه إذا جعلت إحدى عينيك تراقب سلسلة معقدة من أنماط سريعة التغير، سيؤدي ذلك إلى تشتيت انتباه نظامك المرئي إلى الحد الذي يجعلك غير مدرك تماماً لصورة ثابتة معروضة للعين الأخرى.[11] باختصار، يمكنك الحصول على صورة مرئية في شبكيتك دون أن تحس بها، ويمكنك (أثناء الحلم) تجربة صورة دون أن تكون مرئية في شبكيتك. هذا يثبت أن شبكيتيك لا تستضيفان وعيهما البصري في نفسيهما مثلهما مثل كاميرا الفيديو، على الرغم من أنهما تؤديان حوسبات معقدة تضم أكثر من مئة مليون خلية عصبية.



الشكل 4.8: أي مربع أغمق - A أو B؟ ماذا ترى على اليمين - إناء، امرأتان أو كليهما بالتناوب؟ مثل هذه الخدع البصرية تثبت أن وعيك البصري لا يمكن أن يكون في عينيك أو في المراحل المبكرة الأخرى من نظامك البصري، لأنه لا يعتمد فقط على ما يوجد في الصورة.

كما يستخدم باحثو الارتباطات NCC التثبيط المستمر للومضات الخدعَ البصرية/ السمعية غير المستقرة وغيرها من الحيل لتحديد أيّ من مناطق دماغك هي المسؤولة عن كل تجربة من تجاربك الواعية. تتمثل الاستراتيجية الأساسية في مقارنة ما تقوم به الخلايا العصبية في حالتين يكون كل شيء فيهما (بما في ذلك المدخلات الحسية) هو نفسه -باستثناء تجربتك الواعية. وتُعرَّف أجزاء الدماغ التي يتم قياسها على أنها تتصرف بشكل مختلف على أنها ارتباطات NCCs.

أثبتت أبحاث الارتباطات NCC من مثل هذه أن لا شيء من وعيك يكمن في أمعائك، على الرغم من أن هذا هو موقع جهازك العصبي ذي النصف بليون خلية عصبية ضخمة التي تحسب كيفية هضم طعامك على النحو الأمثل؛ غير أن إنتاج مشاعر مثل الجوع والغثيان يحدث في عقلك. وبالمثل، لا يبدو أن أيا من وعيك موجود في جذع الدماغ، وهو الجزء السفلي من الدماغ الذي يتصل بالحبل الشوكي ويتحكم في التنفس ومعدل ضربات القلب وضغط الدم. أما الأكثر إثارة للدهشة: لا يبدو أن وعيك يمتد إلى مخيخك (الشكل 3.8)، الذي يحتوي على نحو ثلثي جميع خلاياك العصبية: المرضى الذين دُمِّر مخيخهم يعانون اضطرابَ النطق والحركة الخرقاء التي تذكرنا بالسُكر. لكنهم ما يزالون واعين تماماً.

وسؤال أيٍّ من الأجزاء المخية هو المسؤول عن الوعي يظل سؤالا مطروحا ومثيرا للجدل. تشير بعض الأبحاث الحديثة التي أجرتها الارتباطات NCC إلى أن وعيك توجد بشكل أساسي في "المنطقة الساخنة" Hot zone التي تشمل المهاد Thalamus (بالقرب من منتصف دماغك) والجزء الخلفي من القشرة (طبقة الدماغ الخارجية التي تتكون من صفيحة مُطعّجة تتألف من ستة طبقات، والتي إذا ما بُسِطت فستعادل مساحتها مساحة منديل عشاء كبير.[12] يشير هذا البحث نفسه بشكل مثير للجدل إلى أن القشرة البصرية الأساسية Primary visual cortex في الجزء الخلفي من الرأس هي استثناء من هذا، كونها لاواعية مثل مقلتي عينيك وشبكيّتيهما.

حتى الآن، نظرنا في الأدلة التجريبية ذات الصلة بأنواع معالجة المعلومات التي تكون واعية، وأين يحدث الوعي. لكن متى يحدث؟ عندما كنت طفلاً، اعتدت على الاعتقاد أننا ندرك الأحداث فور حدوثها، دون تأخير. وعلى الرغم من أن هذا مايزال شعوري الذاتي، فإن من الواضح أن ذلك لا يمكن أن يكون صحيحاً. إذ يستغرق ذهني بعض الوقت لمعالجة المعلومات التي تدخل عبر أعضائي الحسية. قاس باحثو الارتباطات NCC الوقت بدقة، والخلاصة التي توصل إليها كريستوف كوتش هي أن الأمر يستغرق نحو ربع ثانية من الوقت الذي يدخل فيه الضوء إلى عينك من جسم معقّد إلى أن تدرك بوعي رؤيته على ما هو عليه.[13] وهذا يعني أنه إذا كنت تقود سيارتك على طريق سريع بسرعة 50 ميلا في الساعة ورأيت فجأة سنجابا على بعد أمتار قليلة أمامك، يكون قد فات الأوان لأن تفعل أي شيء حيال ذلك، لأنك تكون قد دهسته بالفعل!

باختصار، يعيش وعيك في الماضي، مع تقدير كريستوف كوتش أنه يتخلف عن العالم الخارجي بنحو ربع ثانية. من المثير للاهتمام، أنه يمكنك في كثير من الأحيان أن تتفاعل مع الأشياء بشكل أسرع من إدراكك لها، الأمر الذي يثبت أن معالجة المعلومات المسؤولة عن ردود فعلك الأسرع يجب أن تكون غير واعية. مثلاً، إذا اقترب جسم غريب من عينك، فيمكن أن تغلق عينك الوامضة جفنك في غضون عُشرٍ من الثانية فقط. يبدو الأمر كما لو أن أحد أنظمة دماغك يتلقى المعلومات المُحذِّرة من النظام البصري، ويحسب أن عينك معرضة لخطر الإصابة، فيرسل تعليمات لعضلات عينيك وفي الوقت نفسه يرسل رسالة إلكترونية إلى الجزء الواعي من دماغك قائلاً: «مهلاً، سنرمش». بحلول وقت قراءة هذه الرسالة الإلكترونية وإدراجها في تجربتك واعية، تكون العين قد رمشت بالفعل.

بالطبع، فإن النظام الذي يقرأ تلك الرسالة الإلكترونية يتعرض باستمرار للقصف بوابل من الرسائل من جميع أنحاء جسمك، بعضها متأخر أكثر من غيره. يستغرق وصول الإشارات العصبية إلى عقلك من أصابعك وقتا أطول من تلك التي من وجهك بسبب المسافة، ويستغرق تحليل الصور وقتاً أطول من الأصوات لأنها أكثر تعقيداً - وهذا هو السبب في أن السباقات الأولمبية تبدأ بانفجار صوتي بدلاً من إشارة بصرية. ولكن، إذا لمست أنفك، فإنك تستشعر بوعي الإحساس على أنفك وأطراف أصابعك في الوقت نفسه، وإذا كنت تصفق بيديك، فتسمع وتشعر بالتصفيق في الوقت نفسه.[14] وهذا يعني أن تجربتك الواعية الكاملة بحدث ما لن تحدث حتى وصول آخر رسالة الإلكترونية وتحليلها.

فئة شهيرة من الارتباطات NCC كان رائدها الفسيولوجي بنيامين ليبت Benjamin Libet الذي بيّن أن هذا النوع من الأفعال اللاواعية لا يقتصر فقط على الاستجابات السريعة مثل رمشة العين وضربة كرة تنس الطاولة، ولكنه يشمل أيضاً بعض القرارات التي قد تنسبها إلى الإرادة الحرة - قياسات الدماغ يمكنها في بعض الأحيان أن تتنبأ بقرارك قبل أن تعي أنه اتخذ بالفعل.[15]

نظريات الوعي

رأينا للتو أنه على الرغم من أننا مازلنا لا نفهم الوعي، فلدينا كميات مذهلة من البيانات التجريبية حول جوانبه المختلفة. لكن كل هذه البيانات تأتي من الأدمغة، فكيف يمكن أن يعلمنا هذا أي شيء عن الوعي في الآلات؟ هذا يتطلب استقراء واسعا يقع خارج نطاقنا التجريبي الحالي، وبعبارة أخرى، فإنه يتطلب **نظرية**.

لماذا نظرية؟

لنقدر لماذا نحتاج إلى نظرية حق قدرها، دعنا نقارن نظريات الوعي بنظريات الجاذبية. بدأ العلماء في أخذ نظرية الجاذبية التي وضعها نيوتن على محمل الجد لأنهم استخلصوا منها أكثر مما أدخلوه فيها من بيانات: معادلات بسيطة يمكن أن تكتب على مساحة لا تزيد على مساحة منديل تتنبأ بدقة بنتيجة أي تجربة جاذبية أجريت على الإطلاق. لذلك أخذوا تنبؤاتها على محمل الجد بل إلى ما هو أبعد من المجال الذي اختبرت فيه النظرية، وقد تحولت هذه الاستقراءات الجريئة إلى نجاحات حتى في حركة المجرات في عناقيد تبعد عنا ملايين السنين الضوئية. ولكن، التنبؤات أخطأت بمقدار ضئيل عند تقدير حركة عطارد حول الشمس. ومن ثم، بدأ العلماء يأخذون على محمل الجد نظرية آينشتاين المُحسنَّة للجاذبية، والنسبية العامة، لأنها كانت أكثر أناقة واقتصادية، وتنبؤوا بشكل صحيح حتى بأخطاء نظرية نيوتن. ومن ثم، أخذوا على محمل الجد توقعاتهم وطبقوها أيضاً خارج نطاق المجال الذي تم اختباره فيها، وذلك على ظواهر غرائبية مثل الثقوب السوداء، وموجات الجاذبية (الموجات الثقالية) في الزمكان Spacetime، وتُوسُّع كوننا من الأصل الساخن الملتهب- جميع ذلك أُثبت لاحقاً بالتجربة.

وبالمثل، إذا كانت نظرية رياضياتية للوعي يمكن لمعادلاتها أن تكتب على مساحة لا تزيد على منديل أن تتنبأ بنجاح بنتائج جميع التجارب التي نجريها على العقول، ثم نبدأ في أخذ - ليس فقط النظرية نفسها- ولكن أيضاً تنبؤات للوعي فيما وراء العقول، مثلا في الآلات- على محمل الجد.

الوعي من منظور الفيزياء

على الرغم من أن بعض نظريات الوعي تعود إلى العصور التاريخية القديمة، إلا أن معظم النظريات الحديثة ترتكز على علم النفس العصبي وعلم الأعصاب، في محاولة لشرح وتوقع الوعي من حيث الأحداث العصبية التي تحدث في الدماغ.[16] وعلى الرغم من أن هذه النظريات طرحت عددا من التنبؤات الناجحة حول الارتباطات العصبية للوعي، فإنها لم تطمح إلى أن تطرح توقعات بخصوص وعي الآلات. ولتحقيق القفزة من العقول إلى الآلات، نحتاج إلى التعميم من الارتباطات NCCs إلى الارتباطات PCC: **الارتباطات الفيزيائية**

للوعي ***Physical correlates of consciousness***، مُعرَّفة كأنماط جزيئات متحركة واعية. لأنه إذا تمكنت النظرية من التنبؤ بشكل صحيح بما هو واعٍ وما هو ليس بواعٍ بالرجوع فقط إلى لبنات بناء فيزيائية مثل الجسيمات الأولية وحقول القوة، فيمكنها عندئذٍ طرح تنبؤات ليس للأدمغة فحسب، بل وأيضاً لأي ترتيبات أخرى للمادة، بما في ذلك نظم الذكاء الاصطناعي في المستقبل. لذلك دعونا نأخذ منظور الفيزياء بعين الاعتبار: أيُّ ترتيبات الجسيمات ترتيبات واعية؟

لكن هذا فعليا يثير سؤالاً آخر: كيف يمكن نشوء شيء ما معقد مثل الوعي من شيء بسيط مثل الجسيمات؟ أعتقد أن سبب ذلك هو كونها ظاهرة ذات خصائص تتجاوز خصائص جسيماتها. في الفيزياء، نطلق على مثل هذه الظواهر اسم Emergent "الناشئة".[17] دعونا نحاول فهم هذا من خلال النظر إلى ظاهرة ناشئة أبسط من الوعي: البلل.

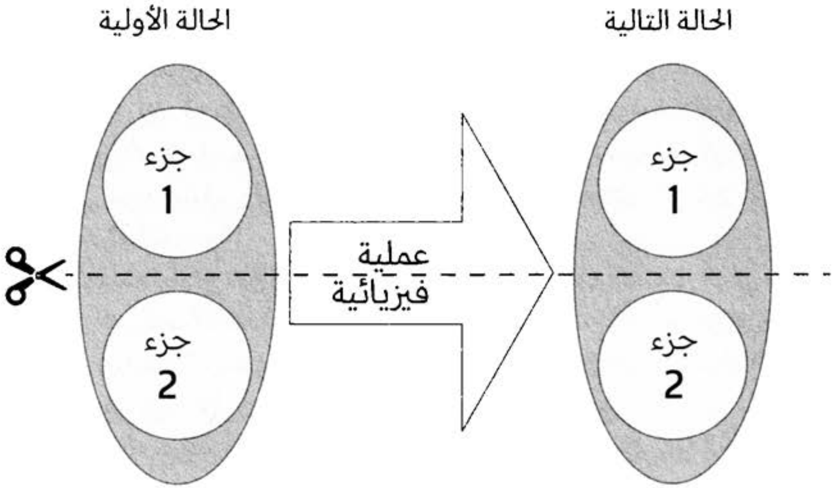
قطرة الماء مبللة. لكن بلورة ثلج وسحابة بخار ليستا كذلك، على الرغم من أنهما تتألفان من جزيئات ماء متماثلة، لماذا؟ لأن خاصية البلل تعتمد فقط على ترتيب الجزيئات. فليس من المنطقي مطلقاً القول إن جزيء الماء الواحد رطب، لأن ظاهرة البلل لا تظهر إلا عند وجود العديد من الجزيئات، مرتبة وفقاً للنمط الذي نسميه سائلاً. لذلك فإن المواد الصلبة والسوائل والغازات كلها ظواهر ناشئة: إنها أكثر من مجرد حاصل مجموع أجزائها، لأنها لها خصائص تتجاوز خصائص جزيئاتها. لديها خصائص تفتقر إليها جزيئاتها.

الآن تماماً مثل المواد الصلبة والسوائل والغازات، أعتقد أن الوعي ظاهرة ناشئة، لها خصائص تتجاوز خصائص جزيئاتها. مثلاً، الدخول في نوم عميق يطفئ الوعي، بمجرد إعادة ترتيب الجزيئات. بالمثل، سيختفي وعيي إذا تجمدت حتى الموت؛ مما يعيد ترتيب جزيئاتي بطريقة مؤسفة.

عندما تضع الكثير من الجزيئات معا لتكوين أي شيء من الماء إلى الدماغ، تبزغ ظواهر جديدة ذات خصائص يمكن رصدها. نحن علماء الفيزياء نحب دراسة هذه الخصائص الناشئة، التي يمكن غالباً تعريفها بمجموعة صغيرة من الأرقام التي يمكنك قياسها - كميات مثل مدى لزوجة المادة، ومدى قابليتها للانضغاط، وهلم جرا. مثلاً، إذا كانت المادة لزجة لدرجة تكون معها غير مرنة، فإننا نسميها مادة صلبة، وإذا كانت على العكس من ذلك فإننا نسميها مادة سائلة. وإذا لم يكن السائل قابلا للانضغاط، فإننا نسميه سائلاً، وإلا فإننا نسميه غازاً أو بلازما، وذلك وفقا لمدى قدرتها على توصيل الكهرباء.

الوعي كمعلومات

إذاً هل يمكن أن يكون هناك كميات مماثلة تقيس الوعي؟ اقترح عالم الأعصاب الإيطالي جوليو تونوني Giulio Tononi إحدى هذه الكميات، التي يسميها **"المعلومات المتكاملة" *Integrated information***، ويرمز إليها بالحرف اليوناني Φ **(تنطق فاي *Phi*)**، الذي يقيس أساساً مدى معرفة أجزاء مختلفة من النظام ببعضها البعض (انظر الشكل 5.8).



Φ تقيس عدم قدرة العملية على الانقسام إلى قسمين مستقلين.

الشكل 5.8: إذا أخذنا عملية فيزيائية بعين الاعتبار فإنها ، مع مرور الوقت، تحول الحالة الأولية للنظام إلى حالة جديدة، تقيس معلوماته المتكاملة Φ عدم القدرة على تقسيم العملية إلى أجزاء مستقلة. إذا كان مستقبل كل حالة مستقبلية لكل جزء يعتمد فقط على ماضيها، وليس على ما يفعله الجزء الآخر، إذاً o = Φ: ما أطلقنا عليه نظاماً واحداً هو في الحقيقة نظامان مستقلان لا يتواصل أحدهما مع الآخر على الإطلاق.

التقيت بجوليو أول مرة في مؤتمر للفيزياء في 2014 ببورتوريكو، الذي دعوته هو وكريستوف كوتش إلى حضوره. وقد شعرت كما لو كان رجلا مثاليا من عصر النهضة يمكنه الاندماج مع غاليليو وليوناردو دافنشي: سلوكه اللطيف لا يخفي معرفته المذهلة بالفن والأدب والفلسفة، وتسبقه سمعته كطاهٍ: مؤخرا أخبرني صحفي تلفزيوني عالمي كيف أعد جوليو في بضع دقائق فقط ألذ سلطة تذوقها في حياته. وسرعان ما أدركت أنه خلف واجهة سلوكه اللطيف كان هناك فكر لا يعرف الخوف يتابع الدليل أينما أخذه، بغض النظر عن الأفكار المسبقة وتابوهات المؤسسة العلمية. تماماً كما اتبع غاليليو نظريته الرياضياتية للحركة على الرغم من ضغوط المؤسسة العلمية لعدم تحدي مركزية الأرض، فقد طوّر جوليو نظرية الوعي الأكثر دقة رياضياتيا حتى الآن، **نظرية المعلومات المتكاملة** ***Integrated information theory*** (اختصارا: النظرية IIT).

لقد جادلت على مدار عقود من الزمن بأن الوعي هو الطريقة التي تشعر بها المعلومات عند معالجتها بطرق معقدة معينة.[18] وتتفق النظرية IIT مع هذا وإن كانت ستستعيض عن عبارة «طرق معقدة معينة» بعبارة محددة: معالجة المعلومات يجب أن تكون جزءا متكاملا، أي Φ يجب أن يكون كبيرا. وحجة جوليو التي يسوقها في هذا المقام قوية بقدر ما هي بسيطة: يحتاج النظام الواعي إلى التكامل في وحدة مكتملة موحدة، لأنه إذا كانت

مكونة، بدلاً من ذلك، من جزأين مستقلين، فسيبدوان وكأنهما كيانان واعيان منفصلان بدلاً من كيان واحد. بعبارة أخرى، إذا لم يتمكن جزء واعٍ من الدماغ أو الحاسوب من التواصل مع الأجزاء الباقية، فلن يكون الباقي جزءاً من تجربته الذاتية.

قاس جوليو ومعاونوه نسخة مبسطة من التخطيط EEG لقياس استجابة الدماغ للتحفيز المغناطيسي. جهازهم «الكاشف للوعي» يعمل بشكل جيد فعلا: لقد حدد أن المرضى كانوا واعين عندما كانوا مستيقظين أو يحلمون، ولكنهم غير واعين عندما تم تخديرهم أو إذا كانوا يغطون في نوم عميق. بل انه اكتشف الوعي في اثنين من المرضى يعانيان متلازمة "المُنْحَبِس" Syndrome Lockedin، واللذين لا يتمكنان من التحرك أو التواصل بأي طريقة طبيعية.[19] لذا فإن هذه التقنية برزت كتكنولوجيا واعدة للأطباء في المستقبل لمساعدتهم على معرفة ما إذا كان بعض المرضى واعين أم لا.

ترسيخ الوعي باستخدام الفيزياء

النظرية IIT مُعرَّفة فقط بالنسبة إلى النظام المنفصل الذي يمكن أن يكون عددا محدودا من الحالات، مثلاً، البتات في ذاكرة الحاسوب، أو الخلايا العصبية المفرطة التبسيط التي قد تكون إما شغالة أو مُطفأة. يعني هذا مع الأسف أن النظرية IIT لم يتم تعريفها بالنسبة إلى معظم الأنظمة الفيزيائية التقليدية، التي يمكن أن تتغير بشكل مستمر - مثلاً، يمكن لموضع الجسيم أو قوة الحقل المغنطيسي أن يأخذ أياً من عدد لا حصر له من القيم.[20] إذا حاولت تطبيق معادلة النظرية IIT على مثل هذه الأنظمة، فستحصل على النتيجة غير المفيدة القائلة إن هو عدد لا منته. فالأنظمة الميكانيكية الكمية منفصلة Quantum-mechanical systems قد تكون أعدادا صحيحة. لكن النظرية IIT الأصلية غير مُعَّرفة للتعامل مع للأنظمة الكمية. فكيف يمكننا أن نقيم النظرية IIT ونظريات الوعي الأخرى القائمة على المعلومات على أساس مادي متين؟

يمكننا القيام بذلك بالاعتماد على ما تعلمناه في الفصل الثاني حول الكيفية التي يمكن بها أن يكون لكتل المادة خصائص طارئة مرتبطة بالمعلومات. لقد رأينا أنه لكي يكون هناك شيء يمكن استخدامه كجهاز ذاكرة يمكنه تخزين المعلومات، فإنه يحتاج إلى العديد من الحالات الطويلة العمر. ورأينا أيضاً أن كونها **كومبيوترونيوم *Computronium***، وهي مادة يمكنها إجراء العمليات الحسابية. إضافة إلى ذلك تتطلب ديناميكيات معقدة: تحتاج قوانين الفيزياء إلى تغييرها بطرق معقدة بما يكفي لتكون قادرة على تنفيذ معالجة المعلومات الاعتباطية. أخيرا رأينا كيف أن الشبكة العصبية، مثلاً، هي مادة أساسية للتعلم لأنه، ببساطة عن طريق إطاعة قوانين الفيزياء، يمكنها إعادة تحديد نطاق نفسها لتصبح أفضل وأفضل في تنفيذ الحوسبات المطلوبة. الآن لنطرح سؤالاً إضافياً: ما الذي يجعل نقطة من المادة قادرة على الحصول على تجربة ذاتية؟ وبعبارة أخرى، في ضوء أي ظروف ستكون نقطة من المادة قادرة على القيام بهذه الأشياء الأربعة؟

1. تذكر
2. حوسبة
3. تعلم
4. شعور بخبرة ما

استكشفنا الثلاثة الأولى في الفصل الثاني. ونحن الآن نعالج الرابعة. تماماً كما صاغ مارغولوس وتوفولي مصطلح **كومبيوترونيوم *Computronium*** لمادة يمكنها أن تؤدي عمليات حوسبية اعتباطية، أود أن أستخدم مصطلح **سينترونيوم *Sentronium*** للمواد الأكثر عمومية التي تتمتع بخبرة ذاتية (قابلة للاحساس أو الوعي).*

ولكن كيف يمكن أن نشعر بالوعي كما لو أنه غير مادي إذا كان في الواقع ظاهرة مادية؟ كيف يمكن أن تشعر أنه مستقل عن مادته الفيزيائية؟ أعتقد أن السبب أنه مستقل عن مادته الفيزيائية، أي عن الأشياء التي يتجلي فيها كنمط! لقد شاهدنا أمثلة جميلة عن أنماط تعتمد على المادة الأساسية في الفصل الثاني. بما في ذلك الموجات والذكريات والحوسبات. لقد رأينا كيف أنها لم تكن مجرد مجموع أجزائها (ناشئة)، بل في الواقع مستقلة عن أجزائها، متخذة مسارا خاصا بها. مثلاً، رأينا كيف لن يُتاح لعقل محاك في المستقبل أو شخصية لعبة حاسوب أي طريقة لمعرفة ما إذا كان يعمل على نظام ويندوز Windows أو أو إس Mac OS أو آندرويد Android أو أي نظام تشغيل آخر، لأنه سيكون مستقلاً عن المادة الأساسية. كما لن تتمكن من معرفة ما إذا كانت البوابات المنطقية Logic gates لجهاز الحاسوب الخاص بها مصنوعة من الترانزستورات أو الدوائر (الدارات) الضوئية Optical circuits أو غيرها من الأجهزة. أو ما هي القواعد الأساسية للفيزياء - يمكن أن تكون أي شيء ما دامت تسمح ببناء أجهزة حاسوب عامة Universal computers.

باختصار، أعتقد أن الوعي ظاهرة مادية تُستشعر كما لو كانت غير مادية لأنها مثل الموجات والحوسبات: ذات خصائص مستقلة عن مادتها الأساسية المُعيَّنة. هذا ينبع منطقيا من فكرة: الوعي كمعلومات. وهذا يؤدي إلى فكرة جذرية أحبها حقاً: إذا كان الوعي هو الطريقة التي تشعر بها المعلومات عند معالجتها بطرق معينة، فيجب أن تكون مستقلة عن المادة الأساسية؛ إذاً فالمهم فقط هو بُنية معالجة المعلومات، وليس بنية المادة التي تُعالج المعلومات، وبعبارة أخرى، فإن الوعي مستقل عن المادة الأساسية لسببين!

كما رأينا، تصف الفيزياء أنماطاً في الزمكان تتطابق مع حركة الجسيمات. إذا كانت ترتيبات الجسيمات تطيع مبادئ معينة، فإنها تؤدي إلى ظهور ظواهر ناشئة مستقلة تماماً عن طبقة

* على الرغم من أنني استخدمت في وقت سابق "بيرسيبتوريوم" (الإدراك الحسي) Perceptronium كمرادف لسينترونيوم (قابل للإحساس أو الوعي) Sentronium، فإن ذلك المصطلح يشير إلى تعريف ضيق للهدف، لأن الإدراك هو مجرد تجارب ذاتية ندركها استناداً إلى المدخلات الحسية – مثلاً، الأحلام والأفكار المولدة داخلياً.

الجسيمات، ومختلفة تماماً عنها. أعظم مثال على ذلك هو معالجة المعلومات، في الكومبيترونيوم. لكننا نقلنا هذه الفكرة الآن إلى مستوى آخر: **إذا كانت معالجة المعلومات نفسها تطيع مبادئ معينة، فقد تؤدي إلى ظهور ظاهرة ناشئة ذات مستوى أعلى نسميها الوعي**. هذا يضع تجربتك الواعية ليس على مستوى واحد بل مستويين، لا عجب أن تشعر بعقلك كما لو أنه غير مادي!

وهذا يثير السؤال التالي: ما هذه المبادئ التي يجب أن تطيعها معالجة المعلومات لتكون واعية؟ لا أدعي معرفة الشروط **الكافية** لمنح الوعي، ولكن هنا أربعة شروط **ضرورية** أراهن عليها واستكشفتها في بحثي:

المبدأ	التعريف
مبدأ المعلومات Information principle	يتمتع النظام الواعي بقدرة كبير على تخزين المعلومات
مبدأ الديناميكيات Dynamics principle	يتمتع النظام الواعي بقدرة كبيرة على معالجة المعلومات
مبدأ الاستقلالية Independence principle	يتمتع النظام الواعي باستقلال كبير عن بقية العالَم
مبدأ التكامل Integration principle	لا يمكن أن يتكون النظام الواعي من أجزاء مستقلة تقريباً

كما قلت، أعتقد أن الوعي هو الطريقة التي تشعر بها المعلومات عند معالجتها بطرق معينة. هذا يعني أنه لكي يكون هناك وعي، يجب أن يكون النظام قادراً على تخزين المعلومات ومعالجتها؛ مما يشير ضمنيا إلى المبدأين الأولين، لاحظ أن الذاكرة لا تحتاج إلى أن تدوم طويلاً: أوصي بمشاهدة الفيديو المؤثر لكلايف ويرثينغ Clive Wearing، الذي يبدو واعياً تماماً على الرغم من أن ذكرياته تدوم أقل من دقيقة.[21] أعتقد أن النظام الواعي يجب أن يكون أيضاً إلى حد ما مستقلاً عن بقية العالم؛ لأنه بخلاف ذلك لن يشعر ذاتياً بأن لديه أي وجود مستقل على الإطلاق. وأخيراً، أعتقد أنه يجب دمج النظام الواعي بشكل متكامل في كيان واحد، كما جادل جوليو تونوني، لأنه إذا تكوّن من جزأين مستقلين، فسيبدو وكأنهما كيانان واعيان منفصلان، وليس كيانا واحدا. تشير المبادئ الثلاثة الأولى ضمنيا إلى الذاتية Autonomy: إن النظام قادر على الاحتفاظ بالمعلومات و بمعالجتها دون الكثير من التدخل الخارجي، ومن ثم تحديد مستقبله بنفسه. المبادئ الأربعة مجتمعة تعني أن النظام مستقل ولكن أجزاءه ليست كذلك.

إذا كانت هذه المبادئ الأربعة صحيحة، فعندئذ ستكون أمامنا مهمة صعبة جدا: سنحتاج إلى البحث عن نظريات راسخة رياضياتيا تجسدها وتختبرها بشكل تجريبي. سنحتاج أيضاً إلى تحديد ما إذا كانت هناك حاجة إلى مبادئ إضافية. بغض النظر عما إذا كان النظرية IIT صحيحة أم لا، يتعين على الباحثين محاولة تطوير نظريات متنافسة واختبار جميع النظريات المتاحة بتجارب أفضل مما طبقت في أي وقت مضى.

الجدل حول الوعي

سبق أن ناقشنا الجدل الأبدي حول ما إذا كانت أبحاث الوعي هراء غير علمي ومضيعة للوقت بلا جدوى. إضافة إلى ذلك، يدور جدل جديد حول أحدث أبحاث الوعي - دعنا نستكشف تلك التي أعتقد أنها أكثر تنويرا.

لم تستقطب نظرية جوليو تونوني -النظرية IIT- مؤخراً الثناء فحسب، بل أثارت النقد أيضاً، وكان بعضه لاذعا. مؤخراً، كتب سكوت آرونسون Scott Aaronson في مدونته: «في رأيي، حقيقة أن نظرية المعلومات المتكاملة خاطئة -خطأ يمكن إيضاحه لأسباب كامنة في جوهرها- تضعها في شريحة الـ 2% العليا من إجمالي النظريات الرياضياتية المقترحة لتفسير الوعي. يبدو لي أن كل نظريات الوعي المتنافسة تقريباً، غامضة للهدف، وغير جادة وقابلة للتطويع، بحيث لا يمكنها إلا أن تطمح إلى الخطأ».[22] وإنصافا لكل من سكوت وجوليو، تمتع كلاهما بضبط النفس عندما شاهدتهما يناقشان النظرية IIT في ورشة عمل عقدت مؤخرا في جامعة نيويورك، واستمعا كل منهما بأدب إلى حجج الآخر. أظهر آرنسون أن شبكات بسيطة معينة من البوابات المنطقية كانت تحتوي على معلومات متكاملة للهدف (Φ) وجادل بأنها لما كانت لاواعية بوضوح، فإن النظرية IIT نظرية خاطئة. ورد جوليو قائلاً إنه إذا تم بناؤها، فستكون واعية، وأن افتراض سكوت كان متحيزاً لكونه مبنياً على الصورة البشرية، كما لو أن مالك المسلخ زعم أن الحيوانات لا يمكن أن تكون واعية لمجرد أنها لا تستطيع التحدث وأنها مختلفة جدّاً عن البشر. كان تحليلي، الذي اتفقا عليه، أنهما كانا على خلاف حول ما إذا كان الاندماج مجرد شرط **ضروري** للوعي (وهو ما وافق عليه سكوت) أو أيضاً شرطا **كافيا** (كما يدعى جوليو). من الواضح أن حالة الشرط الأخيرة هي ادّعاء أقوى وأكثر إثارة للجدل، وآمل أن نتمكن قريباً من اختباره تجريبيا.[23]

هناك ادعاء آخر مثير للجدل في النظرية IIT وهو أنه لا يمكن أن تكون بنية الحواسيب اليوم واعية؛ لأن الطريقة التي تتصل بها بواباتها المنطقية توفر تكاملاً منخفضاً جدّاً.[24] وبعبارة أخرى، إذا قمت بتحميل نفسك إلى روبوت مستقبلي عالي القدرة يحاكي بدقة كل عصبون وكل مشتبك في جهازك العصبي، ومن ثم حتى لو كان هذا المستنسخ الرقمي يبدو كأنه يتحدث ويتصرف بشكل لا يمكن تمييزه عنك، فإن جوليو يدعي أن هذا المستنسخ الرقمي سيكون زومبياً لاواعياً دون تجربة ذاتية - الأمر الذي سيكون مخيباً للآمال إذا حمّلت نفسك في سعي منك لتحقيق خلود ذاتي.* لقد طعن كل من ديفيد تشالمرز David Chalmers وأستاذ الذكاء الاصطناعي موري شاناهان Murray Shanahan

* هناك تناقض محتمل بين هذا الادعاء وفكرة أن الوعي مستقل عن المادة الأساسية، لأنه على الرغم من أن معالجة المعلومات قد تكون مختلفة عند أدنى مستوى، فإنها تتطابق مع بعضها في المستويات العليا حيث يتحدد السلوك.

في هذا الادعاء، وذلك بتخيل ما سيحدث إذا استعضت عن الدوائر العصبية في دماغك بالتدريج بأجهزة رقمية افتراضية تحاكيها بدقة تامة.[25] على الرغم من أن **سلوكك** لن يتأثر بالاستعاضة لأنه من المفترض أن المحاكاة دقيقة تماماً، فستتغير تجربتك من الوعي في البداية إلى اللاوعي في النهاية، وفقاً لجوليو. ولكن كيف سيكون شعورك أثناء ذلك، مع تزايد استعاضة الدوائر العصبية؟ عندما يستعاض عن أجزاء دماغك المسؤولة عن **تجربتك** الواعية للنصف العلوي من مجالك البصري، هل ستلاحظ أن جزءاً من مشهدك البصري قد فقد فجأة. لكنك تعرف بشكل غامض ما الذي كان موجوداً هناك، كما يذكر المرضى الذين يعانون «العمى البصري»؟[26] سيكون هذا مبعثا لقلق عميق، لأنه إذا كان بإمكانك أن تشعر بأي فرق بوعي، فيمكنك أيضاً إخبار أصدقائك بذلك الفرق عندما يسألونك - ومع ذلك، لا يمكن افتراضيا أن يتغير سلوكك. الاحتمال المنطقي الوحيد المتوافق مع الافتراضات هو أنه في الحالة نفسها تماماً التي يختفي فيها أي شيء من وعيك، يتم تغيير رأيك بطريقة غامضة بحيث تجعلك تكذب وتنكر أن تجربتك تغيرت، أو أن تنسى أن الأمور كانت مختلفة.

من ناحية أخرى، يعترف موري شاناهان بأن حجة التدريج البديل نفسها يمكن توجيهها إلى أي نظرية تدعي أن بإمكانك أن تتصرف بوعي دون أن تكون واعياً، لذلك قد تميل إلى استنتاج مفاده أن التمثيل والوعي هما الشيء نفسه، و أن السلوك الذي يمكن رصده خارجياً هو كل ما يهم. ولكن بعد ذلك تكون قد وقعت في فخ التنبؤ بأنك لاواعٍ وأنت تحلم، على الرغم من أنك أذكى من ذلك.

الحجة الثالثة الناقدة لنظرية IIT هي ما إذا كان يمكن إنشاء كيان واعٍ من أجزاء تكون واعية بشكل منفصل. مثلاً، هل يمكن للمجتمع ككل أن يكتسب وعيا دون أن يفقد الأفراد فيه وعيهم؟ هل يمكن أن يكون للدماغ الواعي أجزاء تكون واعية أيضاً من تلقاء نفسها؟ النظرية IIT تتنبأ بأن الجواب "لا" حازمة. ولكن ليس الجميع مقتنعين. مثلا، هناك بعض المرضى المصابين بإصابات تقلل بشدة من التواصل بين شطري دماغها يعانون "متلازمة اليد الغريبة" Alien hand syndrome، إذ يجعل الدماغ الأيمن اليد اليسرى تقوم بأمور يزعم المرضى أنهم لم يتسببوا بها أو يفهموها - أحياناً لدرجة أنهم يستخدمون يدهم الأخرى لكبح يدهم "الغريبة". كيف يمكننا أن نكون على يقين من أنه لا يوجد وعيان منفصلان في رؤوسهم، أحدهما في نصف الكرة الأيمن غير قادر على الكلام والآخر في نصف الكرة الأيسر يُنشئ جميع الكلام ويدعي أنه يتحدث عن النصفين؟ تخيل استخدام تقنية مستقبلية لبناء رابط اتصال مباشر بين عقلين بشريين، وزيادة قدرة هذا الرابط تدريجيّاً حتى يصبح التواصل فعالاً بين العقول كما هو داخلها، هل ستأتي لحظة يختفي فيها الوعيان الفرديان فجأة ويحل محلهما فرد واحد كما تتنبأ به النظرية IIT، أم أن الانتقال سيكون تدريجيّاً بحيث يتعايش الوعي الفردي بشكل ما حتى مع ظهور تجربة مشتركة؟

جدل آخر مدهش هو ما إذا كانت التجارب تدرك مدى وعينا بها، لقد رأينا في وقت سابق أنه على الرغم من أننا نشعر أننا واعون بصرياً بوجود كميات هائلة من المعلومات التي تتضمن الألوان والأشكال والأشياء وكل ما يبدو أمامنا، فقد أظهرت التجارب أنه لا

يمكننا تذكر سوى جزء صغير -إلى حد مفزع- من هذا والإبلاغ عنه.[27] حاول بعض الباحثين حل هذا التناقض من خلال التساؤل عما إذا كان لدينا في بعض الأحيان "وعي دون إمكانية الوصول" Consciousness without access، أي تجربة ذاتية لأشياء معقدة للهدف لا يمكن تخزينها في ذاكرتنا العاملة للاستخدام لاحقاً.[28] مثلاً، عندما تمر بتجربة عمى غير مقصود Inattentional blindness بانصراف انتباهك عن أي شيء في مرأى البصر، فإن هذا لا يعني أنه لم يكن لديك أي تجربة بصرية واعية به، فقط أنه لم يتم تخزينه في ذاكرتك العاملة.[29] هل يجب احتساب ذلك كنسيان أم عمى؟ يرفض باحثون آخرون هذه الفكرة القائلة بأنه لا يمكن الوثوق بما يقول الأفراد أنهم مروا به، ويحذرون من تداعياته. يتخيل موري شاناهان تجربة إكلينيكية يُبلِغ فيها المرضى عن تخفيف الألم تخفيفا كاملا بفضل دواء عجيب جديد، ومع ذلك لا توافق عليه لجنة حكومية: «المرضى يعتقدون فقط أنهم لا يشعرون بالألم. لكن بفضل علم الأعصاب، نحن نعرف أفضل».[30] من ناحية أخرى، كانت هناك حالات لمرضى استيقظوا أثناء عملية جراحة فأعطوا دواء لجعلهم ينسون تلك المحنة. هل يجب أن نثق فيما يبلغون عنه لاحقا من أنهم لم يشعروا بأي ألم؟[31]

كيف يمكن أن يشعر الذكاء الاصطناعي بالوعي؟

إذا كان نظام ذكاء اصطناعي ما واعيا في المستقبل، فما الذي سيختبره ذاتياً؟ هذا هو جوهر "المشكلة الأكثر صعوبة" Even harder problem المتمثلة في الوعي، ويجبرنا على الوصول إلى المستوى الثاني من الصعوبة الموضحة في الشكل 1.8. لا نفتقر حالياً إلى نظرية تجيب عن هذا السؤال فحسب، بل لسنا متأكدين مما إذا كان من الممكن منطقيا الإجابة عنه إجابة تامة. على أي حال، ما هي الإجابة التي ستكون إجابة مرضية؟ كيف تشرح لشخص أعمى منذ الولادة كيف يبدو اللون الأحمر؟

لحسن الحظ، فإن عجزنا الحالي عن تقديم إجابة كاملة لا يمنعنا من تقديم إجابات جزئية. من المحتمل أن تستنتج الكائنات الفضائية الذكية التي تدرس النظام الحسي البشري أن الألوان هي كيفيات محسوسة Qualia يربطها الإحساس بكل نقطة على سطح ثنائي الأبعاد (مجالنا البصري)، في حين أن الأصوات لا نحس بها على أنها ذات موضع مكاني، وأن الآلام هي عبارة عن كيفيات محسوسة يربطها الإحساس بأجزاء مختلفة من جسمنا. وباكتشاف أن شبكية أعيننا لديها ثلاثة أنواع من الخلايا المخروطية الحساسة للضوء، قد تستنتج هذه الكائنات أننا نشعر بثلاثة ألوان أساسية وأن جميع أنواع الألوان الأخرى تنتج عن الجمع بينها. وبقياس الوقت الذي تستغرقه الخلايا العصبية لنقل المعلومات عبر الدماغ، قد تستنتج أننا لا نختبر أكثر من نحو عشرة أفكار أو تصورات واعية في الثانية الواحدة، وأنه عندما نشاهد الأفلام على تلفزيوننا بمعدل 24 إطاراً في الثانية، فإننا لا نشعر بهذه التجربة كسلسلة من الصور الثابتة، ولكن كحركة مستمرة. وبقياس مدى سرعة إفراز الأدرينالين في مجرى الدم وطول المدة المتبقية التي يبقى

فيها قبل أن يتحلل، قد يتنبؤون بأننا نشعر بأن نوبات من الغضب تبدأ في غضون ثوانٍ وتستمر لمدة دقائق.

وبتطبيق حجج مماثلة مستندة إلى قوانين الفيزياء، يمكننا طرح بعض التخمينات المدروسة حول جوانب معينة من كيفية شعور وعي اصطناعي. في البدء، مساحة التجارب المحتملة التي قد يخبرها الذكاء الاصطناعي مساحة ضخمة مقارنة بما يمكن أن نختبره نحن البشر، لدينا فئة واحدة من الكيفيات المحسوسة لكل من حواسنا، ولكن الذكاء الاصطناعي قد يكون لديه مجموعة ضخمة من المستشعرات Sensors والتمثيلات الداخلية للمعلومات، لذلك يجب أن نتجنب الوقوع في مأزق افتراض أن الكون كذكاء الاصطناعي يشبه بالضرورة الكون بشرا.

ثانياً، في الثانية الواحدة, يمكن لوعي اصطناعي بحجم الدماغ أن يمر بملايين التجارب أكثر منّا، لأن الإشارات الكهرومغناطيسية تنتقل بسرعة الضوء - ملايين المرات أسرع من إشارات الخلايا العصبية. ولكن، كلما زاد حجم الذكاء الاصطناعي، كانت أفكاره عالمية وذلك لإتاحة وقت لانتقال المعلومات بين جميع أجزائه، كما رأينا في الفصل الرابع. ومن ثم، فإننا نتوقع أن يكون لدى ذكاء اصطناعي "غايا" Gaia بحجم الأرض نحو عشر تجارب واعية في الثانية الواحدة، مثل الإنسان، ولدى ذكاء اصطناعي بحجم المجرة فكرة عالمية واحد فقط كل 100,000 عام أو نحو ذلك - أي ما لا يزيد عن مئة تجربة خلال تاريخ الكون بكامله حتى الآن! وهذا من شأنه أن يعطي ذكاءً اصطناعيا كبيرا حافزا لا يقاوم لتفويض الحوسبات إلى أصغر النظم الفرعية القادرة على التعامل معها، لتسريع الأمور، إلى حد كبير مثل فوّض عقلنا الواعي رد فعل رمشة العين إلى نظام فرعي صغير وسريع وغير واعٍ. وعلى الرغم من أننا رأينا آنفا أن معالجة المعلومات الواعية في أدمغتنا تبدو وكأنها مجرد الطرف الصغير البارز من جبل جليدي غير واعي بخلاف ذلك الطرف، يجب أن نتوقع أن يكون الوضع أكثر حدّة في الذكاءات الاصطناعية الكبيرة في المستقبل: إذا كان لديها وعي واحد، فمن المحتمل أن أن تكون على علم تقريباً بكل معالجة للمعلومات تجري داخلها. إضافة إلى ذلك، على الرغم من أن التجارب الواعية التي تتمتع بها هذه الذكاءات قد تكون معقدة للهدف، فإنها تتميز أيضاً بأنها تتم بسرعة الحلزون مقارنة بالأنشطة السريعة لأجزائها الصغيرة.

إن هذا يقود إلى النقطة التي يجب عندها أن نجد حلا للجدل المذكور آنفا حول ما إذا كانت أجزاء من كيان واعٍ يمكن أن تكون واعية أيضاً. غير أن النظرية IIT تتنبأ بعكس ذلك؛ مما يعني أنه إذا وجد في المستقبل ذكاء اصطناعي كبير لدرجة فلكية وواعٍ، فإن جميع معالجاته للمعلومات تقريباً تكون غير واعية. هذا يعني أنه إذا حسنت حضارة تتألف من ذكاءات اصطناعية أصغر قدراتها على الاتصال إلى درجة نشوء عقل واعٍ على شكل خلية نحل، فإن أي وعي فردي أسرع من البقية فيها بكثير سينطفئ فجأة. من ناحية أخرى، إذا كان تنبؤ النظرية IIT خاطئاً، يمكن لعقل خلية النحل التعايش مع مجموعة مذهلة من العقول الواعية. في الواقع، قد يتخيل المرء تسلسلا هرميا متداخلا من الوعي على جميع المستويات من المستوى المجهري إلى مستوى الكوزموس.

كما رأينا آنفا، فإن معالجة المعلومات اللاواعية في أدمغتنا البشرية تبدو مرتبطة بطريقة التفكير السريعة التي تحدث دون مجهود يذكر والآلية، التي يسميها علماء النفس "النظام 1" System 1.[32] مثلاً، قد يُبلِغ نظامُ 1 وعيَك أن تحليله المعقد جدّاً للبيانات البصرية المُدخَلة توصل إلى أن صديقك المفضل قد وصل، دون إعطائك أي فكرة عن كيفية حدوث الحوسبات. إذا ثبت أن هذا الرابط بين الأنظمة والوعي رابط صحيح، فسيكون من المغري تعميم هذه المصطلحات على الذكاءات الاصطناعية، ويُشار إلى جميع المهام الروتينية السريعة المفوضة إلى الوحدات الفرعية اللاواعية كنظام الذكاء الاصطناعي 1. أما التفكير الذكاء الاصطناعي العام الذي يحدث دون جهد يذكر والبطيء والمُسيطَر عليه، إذا كان واعياً، فسيكون نظام الذكاء الاصطناعي 2. نحن البشر لدينا أيضاً تجارب واعية تتضمن ما سأطلق عليه "النظام 0": التصور السالب Passive الخام الذي يحدث حتى عندما تجلس دون حركة أو تفكير مراقبا العالم من حولك. يبدو أن تعقيد الأنظمة 0 و 1 و 2 أكثر تعقيداً يتزايد من مستوى إلى آخر، لذلك من المدهش أن المستوى الأوسط هو الذي يبدو لاواعيا. تشرح النظرية IIT هذا بالقول إن المعلومات الحسية الخام في النظام 0 مخزّنة في هياكل دماغية تشبه شبكة ذات تكامل عالٍ للهدف، في حين أن النظام 2 يحتوي على تكامل عالٍ بسبب دوائر التغذية الراجعة، إذ يمكن لجميع المعلومات التي أنت واعٍ بها حالياً أن تؤثر في حالات دماغك المستقبلية. من ناحية أخرى، كان التنبؤ بالشبكة الواعية بالتحديد هو الذي أثار انتقاد سكوت آرنسون -المذكور آنفا- للنظرية IIT. باختصار، في يوم ما إذا تمكنت نظرية ما تحل مشكلة الوعي الصعبة جدّاً من اجتياز مجموعة صارمة من الاختبارات التجريبية ونبدأ في أخذ تنبؤاتها على محمل الجد، فإنها أيضاً ستُضيِّق نطاق الخيارات المتاحة للمشكلة الأكثر صعوبة المتمثلة في ماهية تجربة الذكاءات الاصطناعية الواعية المستقبلية.

تعود بعض جوانب تجربتنا الذاتية بوضوح إلى أصولنا التطورية، مثلا رغباتنا العاطفية المتعلقة بالحفاظ على النفس (الأكل والشرب وتجنب التعرض القتل) والإنجاب. هذا يعني أنه يجب أن يكون من الممكن إنشاء ذكاء اصطناعي لا يمر أبدا بتجربة كيفيات محسوسة مثل الجوع أو العطش أو الخوف أو الرغبة الجنسية. كما رأينا في الفصل الأخير، إذا تمت برمجة ذكاء اصطناعي فائق الذكاء بحيث يكون لديه هدف طموح ما، فمن المحتمل أن يسعى جاهداً إلى الحفاظ على الذات من أجل أن يكون قادرا على تحقيق هذا الهدف. ولكن، إذا كانت جزءاً من مجتمع من الذكاءات الاصطناعية، فقد تفتقر إلى خوفنا الإنساني الشديد من الموت: مادامت تحافظ على نفسها، فكل ما ستخسره هو الذكريات التي تراكمت لديها منذ آخر نسخ احتياطي Backup، مادامت أنها واثقة من أن إمكانية استخدام برمجياتها المنسوخة احتياطيا. إضافة إلى ذلك، من المحتمل أن تقلل القدرة على نسخ المعلومات والبرامج بسهولة ما بين الذكاءات الاصطناعية من الإحساس القوي بالفردية الذي يُميِّز وعينا الإنساني: سيكون هناك تمييز أقل بينك وبيني إذا تمكنا من تقاسم ونسخ جميع ذكرياتنا وقدراتنا بسهولة، لذلك قد تشعر مجموعة من الذكاءات الاصطناعية الموجودة بقرب بعضها البعض مثل كائن حي واحد بعقل على شاكلة خلية نحل.

هل سيشعر الوعي الاصطناعي بأن لديه إرادة حرة؟ لاحظ أنه على الرغم من أن الفلاسفة أمضوا آلاف السنين يتجادلون حول ما إذا كانت **لدينا** نحن إرادة حرة، فإنهم لم يتوصلوا إلى إجماع حتى حول كيفية تحديد السؤال.[33] وأنا أطرح سؤالاً مختلفاً، ويمكن القول إنه من الأسهل معالجته، اسمح لي أن أحاول إقناعك بأن الإجابة هي ببساطة «نعم، سيشعر (ذاتيا) أي صانع قرار واعٍ بأن لديه إرادة حرة، بصرف النظر عما إذا كان بيولوجياً أو اصطناعيا». تقع القرارات على طيف يمتد بين حدين اثنين:

1. أنت تعرف بالضبط لماذا قمت بهذا الاختيار بعينه.
2. ليس لديك أي فكرة عن سبب اختيارك لذلك الخيار، فقد شعرت أنك اخترت عشوائياً بناء على رغبة مفاجئة.

تتمحور مناقشات الإرادة الحرة عادة حول الكفاح من أجل التوفيق بين سلوكنا في اتخاذ القرارات الموجهة نحو الأهداف وقوانين الفيزياء: إذا اخترت واحدا من الشرحين التاليين لما قمت به، إذاً أيها صحيح: **«لقد طلبت مواعدتها لأنها تعجبي حقاً»** أو **«جسيماتي جعلتني أفعل ذلك بالتحرك وفقاً لقوانين الفيزياء»**؟ لكننا رأينا في الفصل الأخير أن كليهما صحيح: ما يمكن أن يبدو كأنه سلوك موجّه نحو الأهداف يمكن أن ينشأ عن قوانين فيزيائية حتمية لا هدف لها. بشكل أكثر تحديداً، عندما يتخذ النظام (عقل أو ذكاء اصطناعي) قراراً من النوع 1، فإنه يحوسب ما يقرره باستخدام بعض الخوارزميات المُحدِّدة Deterministic algorithm، والسبب الذي يجعله يشعر أنه قرر هو أنه في الواقع قرر أثناء إجراء الحوسبة ما يجب فعله. إضافة إلى ذلك، كما أكد Seth Lloyd سيث لويد،[34] هناك فرضية مشهورة في علوم الحاسوب تقول إنه بالنسبة إلى جميع الحوسبات تقريباً، لا توجد طريقة أسرع لتحديد نتائجها من تشغيلها فعليّاً. هذا يعني أنه يستحيل عليك عادةً تحديد ما ستقوم به في ثانية في أقل من ثانية؛ مما يساعد على تعزيز تجربتك بامتلاكك لإرادة حرة. على النقيض من ذلك، عندما يتخذ نظام (عقل أو ذكاء اصطناعي) قراراً من النوع 2، فإنه ببساطة يبرمج عقلَه لاتخاذ قراره بناء على مخرجات بعض النظم الفرعية التي تعمل كمولد لرقم عشوائي. في الأدمغة والحواسيب، يمكن إنتاج أرقام عشوائية الفعالة بسهولة عن طريق تضخيم الضوضاء Noise. وبغض النظر عن النقطة على المدى الواقع بين 1 إلى 2، يشعر كل من الوعي البيولوجي والاصطناعي بأن لديهم إرادة حرة: يشعرون أنهم هم في الحقيقة الذين يقررون ولا يمكنهم التنبؤ بيقين بما سيكون هو القرار حتى ينتهوا من التفكير فيه.

يخبرني البعض أن السببية Causality مُهينة، وأنها تجعل عمليات تفكيرهم عديمة المعنى، وأنها تجعلهم "مجرد" آلات. وأنا أجد مثل هذه السلبية سخيفة وغير مبررة. بادئ ذي بدء، لا يوجد ما هو "مجرد" بخصوص أدمغة البشر، التي في اعتقادي، هي أكثر الأشياء المادية تعقيداً في كوننا المعروف. ثانياً، ما البديل الذي يفضلونه؟ ألا يريدون أن تكون عمليات التفكير الخاصة بهم (الحوسبات التي تقوم بها أدمغتهم) هي التي تتخذ قراراتهم؟

إن تجربتهم الذاتية للإرادة الحرة هي ببساطة الكيفية التي تشعر بها حوسباتهم في الداخل: إنهم لا يعرفون نتائج عمليات الحوسبة إلى أن ينتهوا من إجرائها. وهذا ما يعنيه القول إن الحوسبة هي القرار.

معنى

دعونا ننتهِ بالعودة إلى نقطة البداية لهذا الكتاب: كيف نريد أن يكون مستقبل الحياة؟ لقد رأينا في الفصل السابق كيف أن الثقافات المتنوعة حول العالم تسعى جميعها إلى مستقبل غني بالتجارب الإيجابية، ولكن ذلك الجدل الشائك الرائع ينشأ عند التماس الإجماع حول ما ينبغي اعتباره إيجابياً وكيفية تنفيذ المقايضات بين ما هو جيد للأشكال المختلفة من الحياة. لكن دعونا لا نسمح لتلك الخلافات بتشتيت انتباهنا عن الفيل في الغرفة: لا يمكن أن تكون هناك تجارب إيجابية إذا لم تكن هناك تجارب على الإطلاق، أي إذا لم يكن هناك وعي. بعبارة أخرى، من دون وعي، لا يمكن أن يكون هناك سعادة أو خير أو جمال أو معنى أو غرض - مجرد مضيعة هائلة. هذا يعني أنه عندما يسأل الناس عن معنى الحياة كما لو كانت مهمة الكوزموس منح معنى لوجودنا، فإنهم يفهمون الأمر معكوسا: **ليس كوننا هو ما يعطي معنى للكائنات الواعية، ولكن الكائنات الواعية هي التي تعطي معنى لكوننا**. لذا فالهدف الأول على قائمة رغباتنا للمستقبل ينبغي أن يكون الاحتفاظ (ونأمل بالتوسع فيها) بالوعي البيولوجي و/أو الوعي الاصطناعي في الكوزموس، بدلاً من دفعه نحو الانقراض.

إذا نجحنا في هذا المسعى، فكيف سنشعر نحن البشر حول التعايش مع آلات أكثر ذكاءً؟ هل يزعجك صعود الذكاء الاصطناعي الذي لا يحيد عن مساره، وإذا كان الأمر كذلك، فلماذا؟ في الفصل الثالث، رأينا الكيفية التي ينبغي بها أن يكون من السهل نسبياً على التكنولوجيا المدفوعة بالذكاء الاصطناعي تلبية احتياجاتنا الأساسية مثل الأمن والدخل مادامت الإرادة السياسية متوفرة للقيام بذلك. ولكن، ربما تشعر بالقلق من أن التغذية الجيدة والكسوة والإسكان والترفيه ليست كافية. إذا كنا متأكدين من أن الذكاء الاصطناعي سيعتني بجميع احتياجاتنا ورغباتنا العملية، فهل من الممكن أن نشعر بأننا نفتقر إلى المعنى والهدف في حياتنا، مثل حيوانات حديقة الحيوانات المصونة جيداً؟

تقليدياً، أسسنا نحن البشر في كثير من الأحيان قيمتنا الذاتية على **فكرة التفوق الإنساني** ***Human exceptionalism***: القناعة بأننا أذكى الكيانات على هذا الكوكب، ومن ثم فنحن كائنات فريدة ومتفوقة. وظهور الذكاء الاصطناعي سيجبرنا على التخلي عن هذا الأمر وأن نغدو أكثر تواضعاً. لكن ربما هذا شيء يجب أن نفعله على أي حال: فإن التشبث بغطرسة مفاهيم التفوق على الآخرين (الأفراد والجماعات العرقية والأنواع الحية وما إلى ذلك) قد تسبب في مشكلات فظيعة في الماضي، وقد تكون فكرة جاهزة لإحالتها إلى التقاعد. في الواقع، لم تسبب فكرة تفوق البشر الحزن فقط في الماضي، ولكن يبدو أيضاً أنها لا لزوم لها

من أجل ازدهار إنسانيتنا: إذا اكتشفنا حضارة مسالمة خارج كوكب الأرض أكثر تقدماً منا في العلوم والفن وكل شيء آخر يهمنا، فمن المفروض أن هذا لن يمنع الأفراد من الاستمرار في الشعور بوجود معنى وغرض لحياتهم. سنحتفظ بعائلاتنا وأصدقائنا ومجتمعاتنا الأوسع نطاقاً، وجميع الأنشطة التي تمنحنا المعنى والهدف، ونأمل ألا نفقد شيئاً سوى الغطرسة.

ونحن نخطط لمستقبلنا، لننظر في المعنى ليس فقط لحياتنا، ولكن أيضاً للكون نفسه. في هذا المقام، يقدم اثنان من علماء الفيزياء المفضلين لدي، ستيفن واينبرغ Steven Weinberg وفريمان دايسون Freeman Dyson، وجهتي نظر متناقضتين تماماً. واينبرغ، الحاصل على جائزة نوبل لوضع أساس النموذج القياسي لفيزياء الجسيمات Standard model of particle physics، قال عبارته الشهيرة: «كلما بدا أن الكون صار مفهوما أكثر، بدا أيضاً أنه بلا هدف».[35] أما دايسون، من ناحية أخرى، فهو أكثر تفاؤلاً، كما رأينا في الفصل السادس: على الرغم من أنه يوافق على أن كوننا كان بلا معنى، فإنه يعتقد أن الحياة صارت ذات معنى أكثر من أي وقت مضى، والأفضل آتٍ إذا نجحت الحياة في الانتشار في جميع أنحاء الكوزموس. وقد أنهى بحثه المنشور عام 1979 على هذا النحو: «هل عالم واينبرغ أو عالمي أقرب إلى الحقيقة؟ في يوم ما، وقبل مرور فترة طويلة، سنعرف الإجابة».[36] إذا عاد الكون إلى الكون غير واعٍ بشكل دائم لأننا دفعنا الحياة على الأرض إلى الانقراض أو لأننا سمحنا لذكاء اصطناعي زومبي بتولي زمام كوننا، فسيكون واينبرغ على صواب دون منازع.

من هذا المنظور، نرى أنه على الرغم من أننا ركزنا على مستقبل الذكاء في هذا الكتاب، فإن مستقبل الوعي أكثر أهمية، لأن هذا هو ما يُمكّن المعنى. **يحب الفلاسفة اللجوء إلى اللاتينية** في هذا التمييز، من خلال **مقارنة الحكمة** ***Sapience*** (القدرة على التفكير بذكاء) بالإحساس Sentience (القدرة على تجربة الكيفيات المحسوسة تجربة ذاتيا). نحن البشر بنينا هويتنا على أننا نكون (الإنسان العاقل) هومو سابينس Homo sapiens، أذكى الكيانات الموجودة. ونحن آخذين بالاستعداد للتواضع أمام آلات أكثر ذكاء من أي وقت مضى، أقترح أن نعيد تسمية أنفسنا بوصفنا (الإنسان الحساس) **هومو سينتينس** ***Homo sentiens***!

الخلاصة:

- لا يوجد تعريف غير متنازع عليه لـ "الوعي". أستخدم التعريف الواسع وغير المتمركز حول الإنسان **للوعي** ***Consciousness*** **= تجربة ذاتية** ***Subjective experience***.
- ما إذا كانت الذكاءات الاصطناعية واعية بهذا المعنى هو ما يهم بالنسبة إلى المشكلات الأخلاقية والفلسفية الشائكة التي يطرحها صعود الذكاء الاصطناعي: هل يمكن أن يُعاني الذكاءُ الاصطناعي؟ هل يجب أن تكون له حقوق؟ هل تحميل الذات Uploading انتحار شخصي؟ هل يمكن لكوزموس مستقبلي يعج بالذكاء الاصطناعي أن يكون نهاية العالَم بسيادة حالة الزومبي عليه؟

- يجب ألا نخلط بين مشكلة فهم الذكاء والمشكلات الثلاث المنفصلة للوعي: "المشكلة الصعبة جدّاً" للتنبؤ بالأنظمة المادية الواعية، و"المشكلة الأكثر صعوبة" للتنبؤ بالكيفيات المحسوسة Qualia، و"المشكلة الصعبة **حقاً**" لماذا هناك وعي على الإطلاق.
- "المشكلة الصعبة جدّاً" للوعي هي موضوع علمي، لما كانت النظرية التي تتنبأ بأي من عمليات دماغك عمليات واعية يمكن اختبارها ودحضها، فإنه ليس من الواضح بعد الكيفية التي سيتمكن بها العلم من حل المشكلتين الأصعب.
- تشير تجارب علم الأعصاب إلى أن العديد من السلوكيات ومناطق الدماغ غير واعية، إذ إن الكثير من تجربتنا الواعية يمثل ملخصاً لما بعد الواقع لكميات ضخمة من معلومات غير واعية.
- تعميم تنبؤات الوعي من العقول إلى الآلات يتطلب نظرية. ويبدو أن الوعي لا يتطلب نوعاً معيناً من الجسيمات أو الحقول، بل يتطلب نوعاً معيناً من المعلومات التي تتمتّع بالاستقلالية والتكامل إلى حد ما، بحيث يكون النظام بكامله مستقلاً إلى حد ما في حين أن أجزاءه ليست كذلك.
- قد يثير الوعي الشعور بأنه غير مادي مستقل عن المادة استقلالا مضاعفا: إذا كان الوعي هو الطريقة التي تشعر بها المعلومات عندما يتم دمجها بطرق معقدة معينة، فإن بنية معالجة المعلومات هي المهمة فقط، وليست بنية المادة التي تقوم بمعالجة المعلومات.
- إذا كان الوعي الاصطناعي ممكناً، فقد تكون مساحة التجارب التي يخبرها الذكاء الاصطناعي ضخمة مقارنة بما يمكننا نحن البشر أن نختبره، إذ تمتد هذه المساحة عبر طائفة واسعة من الكيفيات المحسوسة والمقاييس الزمنية- يتشارك جميعها في إحساسها بأنها تمتلك إرادة حرة.
- نظراً لأنه لا يمكن أن يكون هناك معنى من دون وعيه، فليس كوننا هو ما يعطي معنى للكائنات الواعية، بل الكائنات الواعية هي التي تسبغ معنىً على كوننا.
- يقترح هذا أننا بينما نستعد نحن البشر للتواضع أمام الأجهزة الذكية الآخذة في التزايد في الذكاء، فإننا سنشعر بالسلوى إذا نظرنا إلى أنفسنا كـ **(الإنسان الحساس) هومو سينتينس *Homo sentiens*، وليس (الإنسان العاقل) هومو سابينس *Homo sapiens***.

الخاتمة

—

حكاية الفريق FLI

الجانب الأكثر تعاسة في الحياة الآن هو أن العلم يجمع المعرفة بشكل أسرع من جمع المجتمع للحكمة.

إسحاق أسيموف

ها نحن، أيها القارئ العزيز، نصل إلى نهاية الكتاب، بعد استكشاف أصل الذكاء ومصيره، والأهداف والمعنى. إذاً كيف لنا أن نترجم هذه الأفكار إلى أفعال؟ وما الأعمال الملموسة التي يجب علينا **القيام** بها لجعل مستقبلنا جيداً قدر الإمكان؟ هذا هو بالضبط السؤال الذي أطرحه على نفسي الآن إذ أجلس هنا في مقعدي عند النافذة في طريقي من سان فرانسيسكو إلى بوسطن في 9 يناير 2017، عائدا من مؤتمر الذكاء الاصطناعي الذي نظّمتُه للتو في أسيلومار، لذلك اسمح لي أن أنهي هذا الكتاب بتقاسم أفكاري معك.

إلى جواري تعوّض ميا ما فاتها من النوم في الليالي العديدة أثناء الإعداد والتنظيم، ياله من أسبوع جامح! فقد تمكنا من حشد جميع الأشخاص الذين ذكرتهم في هذا الكتاب تقريباً لبضعة أيام في هذا الملتقى الذي يأتي تكملة لمؤتمر بورتوريكو، بما في ذلك روّاد مثل إيلون ماسك ولاري بيج وقادة أبحاث الذكاء الاصطناعي من الأوساط الأكاديمية والشركات مثل ديب مايند DeepMind وغوغل Google وفيسبوك FaceBook وآبل Apple وآي بي إم IBM ومايكروسوفت Microsoft وبايدو Baidu، فضلاً عن الاقتصاديين وعلماء القانون والفلاسفة وغيرهم من المفكرين المذهلين (انظر الشكل 1.9). هذا وقد فاقت النتائج توقعاتي العالية، وأشعر بالتفاؤل بشأن مستقبل الحياة لدرجة أكبر مما شعرت به منذ فترة طويلة. سأخبرك عن السبب في هذه الخاتمة.



الشكل 1.9: مؤتمرنا في أسيلومار في يناير 2017، وهو مؤتمر لاحق لمؤتمر بورتوريكو، جمع مجموعة رائعة من الباحثين في الذكاء الاصطناعي والمجالات ذات الصلة. الصف الخلفي، من اليسار إلى اليمين: باتريك لين، دانييل ويلد، أرييل كون، نانسي تشانغ، توم ميتشل، راي كورزويل، دانييل ديوي، مارغريت بودين، بيتر نورفيغ، نيك هاي، موشيه فاردي، سكوت سيسكيند، نيك بوستروم، فرانشيسكا روسي، شين ليغ، مانويلا فيلوسو، ديفيد ماربل، كاتيا غريس، إراكلي بيريدز، مارتي تيننباوم، غيل برات، مارتين ريس، جوشوا غرين، مات شرير، أنجيلا كين، أمارا أنجيليكا، جيف مور، مصطفى سليمان، ستيف أوموهوندرو، كيت كروفورد، فيتاليك بوتين، يوتاكا ماتسو، ستيفانو إيرمون، مايكل ويلمان، باس ستونبرينك، ويندل والاك، ألان دافوي، توبي أورد، توماس ديتريتش، دانييل كانيمان، داريو أمودي، إريك دريكسلر، توماسو بوجيو، إريك شميدت، بيدرو أورتيغا، ديفيد ليكي، شين أو هيغيراتيغ، أوين إيفانز، يان تالين، أنكا دراغان، شون ليغسيك، توبي والش، بيتر أسارو، كاي فيرث-بترفيلد، فيليب سابس، بول ميرولا، بارت سيلمان، تاكر ديفي، جاكو شتاينهاردت، موشيه لوكس، جوش تيننباوم، توم غروبر، أندرو نغ، كريم أيوب، كريغ كالهون، بيرسي ليانغ، هيلين تونر، ديفيد تشالمرز، ريتشارد سوتون، كلوديا باسوس-فيريرا، جانوس كرمار، ويليام ماك أسكيل، اليعازر يودكووسكي، برايان زيبارت، هوو برايس، كارل شولمان، نيل لورنس، ريتشارد ملاح، يورغن شميدهوبر، ديليب جورج، جوناثان روثبرغ، نوح روثبرغ. الصف الأمامي: أنتوني أغيري، سونيا ساكس، لوكاس بيري، جيفري ساكس، فنسنت كونيتزر، ستيف غوس، فيكتوريا كراكوفنا، أوين كوتون-بارات، دانييلا روس، ديلان هادفيلد-مينيل، فيريتي هاردينغ، شيفون زيليس، لوران أورسو، رامانا كومار، نيت سواريس، أندرو مكافي، جاك كلارك، آنا سالامون، لونغ أويانغ، أندرو كريتش، بول كريستيانو، جوشوا بينغيو، ديفيد سانفورد، كاثرين أولسون، جيسيكا تايلور، مارتينا كونز، كريستين ثوريسون، ستيوارت أرمسترونغ، يان لوكون، ألكسندر تاماس، رومان يامبولسكي، مارين سولياسيس، لورانس كراوس، ستيوارت راسل، إريك برينجولفسون، ريان كالو، شاولان هوسيس، ميا تشيتا-تيغمارك، كنت ووكر، هيذر روف، ميريديث ويتاكر، ماكس تيغمارك، أدريان ويلر، خوسيه هيرنانديز-أورلو، أندرو مينارد، جوم هيرينغ، إبراهيم دمشقي، نيكولاس بيرغرين، غريغوري بونيت، سام هاريس، تيم هوانغ، أندرو سنايدر-بيتي، مارتا هالينا، سيباستيان فاركوهار، ستيفن كيف، جان ليك، تاشا مكولي، جوزيف غوردون-ليفيت. وصل في وقت لاحق: غورو بانافار، ديميس هاسابيس، راو كامبهامباتي، إيلون ماسك، لاري بيج، أنتوني روميرو.



ولادة FLI

منذ أن علمت بسباق التسلح النووي في سن الرابعة عشرة، شعرت بالقلق من أن قوة تقنيتنا تنمو بشكل أسرع من الحكمة التي نديرها بها. لذلك قررت أن أُدرج خلسة فصلاً حول هذا التحدي في كتابي الأول **كوننا الرياضياتي** ***Our Mathematical Universe***، على الرغم من أن بقية الكتاب كانت تتعلق بالفيزياء في المقام الأول. وقراري بمناسبة العام الجديد لعام 2014 كان أنني لن أسمح لنفسي بالتذمر من أي شيء دون التفكير بشكل جدي فيما يمكنني فعله شخصياً حيال ذلك، وقد التزمت بالعهد الذي قطعته خلال جولة كتابي في شهر يناير ذلك: وبالكثير من العصف الذهني تناقشنا أنا وميا حول بدء منظمة غير ربحية تركز على تحسين مستقبل الحياة من خلال الريادة التكنولوجية.

وأصرت على أن نطلق عليها اسماً إيجابياً ومختلفا تماماً عن "معهد الدمار والكآبة" Doom & Gloom Institute و"معهد دعنا نقلق بخصوص المستقبل" -Let's-Worry about-the- Future Institute. ولما كان اسم معهد مستقبل البشرية Future of Humanity Institute قد أُخذ بالفعل، فقد اتفقنا على معهد مستقبل الحياة Future of Life Institute (اختصارا: المعهد FLI)، ويمتاز الاسم بميزة إضافية تتمثل بكونه أكثر شمولية. في 22 يناير، قادتنا جولة الكتاب إلى سانتا كروز، ومع غروب الشمس في كاليفورنيا، استمتعنا بالعشاء مع صديقنا القديم أنتوني أغيري Anthony Aguirre وأقنعناه بالانضمام إلينا. هو ليس فقط واحداً من أكثر الأشخاص الذين أعرفهم حكمة ومثالية، ولكنه أيضاً شخص نجح -معي- في إدارة منظمة غير ربحية أخرى، **معهد الأسئلة التأسيسية** (انظر: http://fqxi.org)، لأكثر من عقد من الزمن.

في الأسبوع التالي، أخذتني الجولة إلى لندن. ولما كان مستقبل الذكاء الاصطناعي يشغل خاطري، فقد تواصلت مع ديميس هاسابيس Demis Hassabis، الذي وجه لي دعوة كريمة لزيارة مقر ديب DeepMind. وذهلت من مدى نموهم منذ أن زارني في المعهد MIT قبل عامين. فقد استحوذت غوغل Google للتو عليهم واشترتهم نحو 650 مليون دولار، وتملكني شعور غامر بأن النجاح كان احتمالاً حقيقياً عندما رأيت مساحة مكتبهم الشاسعة ممتلئة بالعقول المذهلة التي تسعى إلى تحقيق هدف ديميس الجريء المتمثل في "حل الذكاء" Solve intelligence.

في المساء التالي، تحدثت مع صديقي يان تالين Jaan Tallinn عبر سكايب Skype، البرنامج الذي ساعد على تطويره. وشرحت له رؤيتنا للمعهد FLI، وبعدها بساعة، قرر أن يراهن علينا بتمويلنا بحد أقصى 100,000 دولار في السنة! قليلة هي الأشياء التي تستثير مشاعري أكثر عندما يثق بي شخص بما هو أكثر مما أستحقه، لذلك بعد عام من مؤتمر بورتوريكو الذي ذكرته في الفصل الأول. قال مازحا إن هذا كان أفضل استثمار قام به على الإطلاق، وعنى قوله هذا لي الكثير.

في اليوم التالي، ترك ناشري وقت فراغ في جدول أعمالي، ملأتها بزيارة متحف العلوم في لندن London Science Museum. ولما كنت قد قضيت فترة طويلة مشغولا بهواجسي بماضي الذكاء الاصطناعي ومستقبله، شعرت فجأة أنني أسير عبر تجسيد مادي لأفكاري. فالمتحف يعرض مجموعة رائعة من الأشياء التي تمثل نمو معرفتنا، من قاطرة روكيت لستيفنسون Stephenson's Rocket locomotive إلى سيارة فورد موديل تي Model T Ford، ونسخة بالحجم الطبيعي من مركبة أبولو 11 القمرية، وأجهزة حاسوب يرجع تاريخها من الآلة الحاسبة الميكانيكية لبابيج Babbage -"آلة الفروق" Difference Engine- إلى الأجهزة الحالية. كما كان لديهم معروضات تلخص تاريخ فهمنا للعقل، من تجارب ساق الضفدع لغالفانو Galvano إلى الخلايا العصبية (العصبونات)، والتخطيط EEG والتصوير fMRI.

نادراً ما أبكي. لكن هذا ما فعلته في طريق عودتي- وفي نفق ممتلىء بالمارة لا أقل، في طريقي إلى محطة مترو أنفاق ساوث كينسينغتون. ها هنا جميع هؤلاء الناس يمضون في حياتهم بسعادة غير مدركين ما كنت أفكر فيه. أولاً، اكتشفنا نحن البشر كيفية نسخ بعض العمليات الطبيعية باستخدام الآلات، فولَّدنا رياحنا وبرقنا، وقدرة أحصنتنا الميكانيكية. تدريجيّاً، بدأنا ندرك أن أجسادنا كانت أيضاً آلات. ثم بدأ اكتشاف الخلايا العصبية في طمس الحدود بين الجسم والعقل. بعدها بدأنا في بناء آلات يمكنها أن تتفوق ليس فقط على عضلاتنا، ولكن عقولنا كذلك. إذاً بالتوازي مع اكتشاف من نحن، هل من المحتم أننا نجعل أنفسنا غرضا عفى عليه الزمن؟ سيكون ذلك حدثا مأساوياً بالتأكيد.

لقد أفزعتني هذه الفكرة. لكنها عززت أيضاً تصميمي على الحفاظ على قراري للسنة الجديدة. شعرت أننا بحاجة إلى شخص آخر لإكمال فريقنا من مؤسسي المعهد FLI، شخص مؤهل لقيادة فريق من المتطوعين الشباب المثاليين. كان الاختيار المنطقي هو

فيكتوريا كراكوفنا Viktoriya Krakovna، وهي طالبة بارعة من جامعة هارفارد لم تفز فقط بالميدالية الفضية في أولمبياد الرياضيات الدولي فحسب، بل إنه أيضاً هي مؤسسة سيتيدال (القلعة) Citadel، منزل لنحو دستة من الشباب المثاليين الذين يبحثون عن سبب يدفعهم لأداء دور أكبر في حياتهم والعالم. دعوناها أنا وميا إلى منزلنا بعد خمسة أيام لإخبارها عن رؤيتنا، وقبل أن ننتهي من أكل السوشي، وُلد المعهد FLI.

مغامرة بورتوريكو

وهذا مثّل بداية مغامرة مذهلة لا تزال مستمرة. كما ذكرت في الفصل الأول، فقد عقدنا اجتماعات منتظمة لتبادل الأفكار في منزلنا مع العشرات من الطلبة والأساتذة وغيرهم من المفكرين المحليين، فتحولت الأفكار الأعلى تصنيفاً إلى مشاريع - أولها المقالة الافتتاحية حول الذكاء الاصطناعي التي شارك في كتابتها ستيفن هوكينغ وستيوارت راسل وفرانك ويلكزك Frank Wilczek وساعدت على إشعال النقاش العام. بالتوازي مع الخطوات الأولى لإنشاء مؤسسة جديدة (مثل إضافة لجنة استشارية واختيار أعضائها وإطلاق موقع إلكتروني)، نظّمنا حدث إطلاقٍ مرحٍ في قاعة في المعهد MIT مزدحمة بالحضور، استكشف خلالها ألان ألدا Alan Alda مع عدد من مع كبار الخبراء مستقبل التكنولوجيا.



شكل 2.9: يان تالين، أنتوني أغيري، المخلص لكم (أنا)، ميا تشيتا-تيغمارك، وفيكتوريا كراكوفنا، محتفلين بتأسيسنا للمعهد FLI مع السوشي في 23 مايو 2014.

ولبقية العام ركزّنا على التحضير لمؤتمر بورتوريكو الذي، كما ذكرت في الفصل الأول. كان يهدف إلى إشراك كبار باحثي الذكاء الاصطناعي العالميين في مناقشة كيفية الحفاظ على الذكاء الاصطناعي مفيدا. كان هدفنا هو تحويل الحوار الدائر حول

السلامة في الذكاء الاصطناعي من القلق إلى العمل: من المشاحنات حول مدى القلق، إلى الاتفاق على مشاريع بحثية ملموسة يمكن أن تبدأ على الفور لتحقيق أقصى ما يمكن من نتائج جيدة. للاستعداد، جمعنا أفكاراً واعدة حول أبحاث السلامة في مجال الذكاء الاصطناعي من جميع أنحاء العالم، وطلبنا تعليقات المجتمع العلمي المعني بالذكاء الاصطناعي ككل على قائمة مشاريعنا المتنامية. وبمساعدة ستيوارت راسل ومجموعة من المتطوعين الشباب المجتهدين، وخاصة دانيال ديوي Daniel Dewey ويانوس كريمار János Krámar وريتشارد ملاح Richard Mallah، لخصنا أولويات البحث هذه في وثيقة ستُناقش في المؤتمر.[1] وكنا نأمل بأن حشد التوافق على وجوب إجراء الكثير من الأبحاث القيمة التي تتناول موضوعات السلامة في الذكاء الاصطناعي سيشجع الأفراد على البدء بمثل هذه الأبحاث. وسيكون الانتصار النهائي إذا ما تمكن ذلك من إقناع شخص ما بتمويله لأنه، حتى ذلك الوقت، لم يكن هناك أي دعم لهذا العمل من قبل وكالات التمويل الحكومية.

وهنا يدخل إيلون ماسك المشهد. في 2 أغسطس، ظهر على رادارنا بتغريدته الشهيرة «يستحق القراءة الذكاء الفائق من تأليف بوستروم. نحتاج إلى أن نكون حذرين جدّاً مع الذكاء الاصطناعي. قد تكون أكثر خطورة من الأسلحة النووية». وتواصلت معه حول جهودنا، وتحدثت معه عبر الهاتف بعد بضعة أسابيع. وعلى الرغم من أنني شعرت بالتوتر نظرا لانبهاري بهذه الشخصية المشهورة، فإن النتيجة كانت رائعة: وافق على الانضمام إلى المجلس الاستشاري العلمي للمعهد FLI وحضور مؤتمرنا واحتمال تمويل برنامج بحثي أول حول سلامة الذكاء الاصطناعي سيعلن عنه في بورتوريكو. وأشعل هذا حماسنا في المعهد FLI، وجعلنا نكثّف جهودنا لإعداد مؤتمر رائع وتحديد موضوعات البحث الواعدة وحشد دعم المجتمع لهذه الموضوعات.

وأخيرا التقيت مع إيلون شخصياً للمزيد من التخطيط عندما زار المعهد MIT بعد شهرين لحضور ندوة فضائية. شعرت بالغرابة في أن أكون وحدي معه في غرفة خضراء صغيرة بعد لحظات فقط من استحواذه على اهتمام أكثر من ألف طالب من طلبة المعهد MIT كما لو كان نجماً من نجوم موسيقى الروك، ولكن بعد بضع دقائق، كان كل ما يمكنني التفكير فيه هو مشروعنا المشترك. راق لي على الفور. فهو يُشع صدقاً، وألهمني مدى اهتمامه بمستقبل إنسانية على المدى الطويل، وكيف حوّل طموحه الجريء إلى أفعال. لقد أراد أن يستكشف الإنسان كوننا ويستوطنه، لذلك أنشأ شركة فضاء. أراد طاقة مستدامة، لذلك أسس شركة للطاقة الشمسية وشركة للسيارات الكهربائية. طويل القامة، وسيم، بليغ وعلى اطلاع بشكل لا يصدق، كان من السهل أن نفهم لماذا يستمع الناس إليه.

مع الأسف، تعلمت أيضاً من هذا الحدث بمعهد MIT الحذر من وسائل الإعلام التي تبعث على الخوف والانقسام. تألفت محاضرة إيلون على المسرح من ساعة من النقاش الرائع حول استكشاف الفضاء، الذي أعتقد أن المحاضرة كانت ستشكل إنتاجا تلفزيونيا رائعا. في النهاية، سأله أحد الطلبة سؤالا خارج الموضوع عن الذكاء الاصطناعي، وتضمنت

إجابته عبارة «بذكاء اصطناعي، نحن نستحضر الشيطان»، وأصبحت هذه العبارة الشيء الوحيد الذي تناقلته معظم وسائل الإعلام - وخارج السياق بشكل عام. لقد أدهشني أن العديد من الصحفيين كانوا يقومون عن غير قصد بالعكس تماماً مما كنا نحاول إنجازه في بورتوريكو. بينما أردنا تعزيز إجماع المجتمع من خلال تسليط الضوء على أرضية مشتركة، كان لدى وسائل الإعلام حافز لتسليط الضوء على الانقسامات. كلما زاد الجدل الذي يمكن أن يقدموه، زادت درجاتهم على مقياس نيلسن Nielsen ratings وإيرادات الإعلانات. إضافة إلى ذلك، بينما كنا نرغب في مساعدة الناس من مختلف أنحاء الآراء على الاجتماع والتواصل مع الآخرين وفهمهم بشكل أفضل، تسببت التغطية الإعلامية عن غير قصد في إنزعاج الناس من بعضهم البعض عبر الطيف الممتد من الآراء؛ مما زاد من سوء التفاهم من خلال نشر عباراتهم الأكثر استفزازية والخارجة عن السياق الأعم. لهذا السبب، قررنا حظر الصحفيين من اجتماع بورتوريكو وفرض "قاعدة تشاتام هاوس" Chatham House Rule التي تحظر على المشاركين الكشف لاحقاً عمّن قال ماذا.*

على الرغم من أن مؤتمرنا في بورتوريكو حقق نجاحاً جيدا، فإنه لم يكن سهلاً. احتاج العد التنازلي في الغالب إلى عمل تحضيري دؤوب، مثلاً، اتصالي بالهاتف أو سكايب بعدد كبير من الباحثين في الذكاء الاصطناعي لحشد كتلة حرجة Critical mass من المشاركين لجذب مشاركين آخرين، وكانت هناك أيضاً لحظات دراماتيكية - مثل عندما استيقظت في الساعة السابعة صباحاً، في 27 ديسمبر للتحدث إلى إيلون في أوروغواي عبر اتصال هاتفي رديء الشبكة، وليقول لي «لا أعتقد أن هذا سيكون ناجحاً». فقد كان قلقا من أن برنامج أبحاث السلامة الذكاء الاصطناعي قد توفر شعوراً زائفاً بالأمان؛ مما يتيح للباحثين المتهورين المضي قدما فيما يعربون لفظيا فقط بالالتزام بقواعد السلامة. ولكن بعد ذلك، على الرغم من تقطّع الصوت باستمرار، ناقشنا بشكل مفصل الفوائد الضخمة المتمثلة في إدراج الموضوع في التيار العام ودفع المزيد من الباحثين في مجال الذكاء الاصطناعي للعمل في مجال سلامة الذكاء الاصطناعي. بعد أن انقطعت المكالمة، أرسل لي إحدى رسائل البريد الإلكتروني المفضلة لدي على الإطلاق: «انقطعت المكالمة على هذا الطرف هنا. على أي حال، تبدو المستندات جديدة، أنا سعيد بدعم البحث بمبلغ خمسة ملايين دولار لمدة ثلاث سنوات، ربما يجب أن نجعلها عشرة ملايين دولار؟»

بعد أربعة أيام، بدأ عام 2015 بداية جيدة لميا ولي عندما ارتحنا قليلا قبل المؤتمر، ونحن نرقص في العام الجديد على شاطئ في بورتوريكو مضاء بالألعاب النارية. وبدأ المؤتمر بداية رائعة أيضاً: كان هناك إجماع ملحوظ على الحاجة إلى مزيد من أبحاث سلامة الذكاء الاصطناعي، وبناء على المزيد من المداخلات من المشاركين في المؤتمر، تم تحسين وثيقة

* هذه التجربة دفعتني أيضاً إلى إعادة التفكير في الطريقة التي يجب أن أفسر بها الأخبار شخصياً. على الرغم من أنه من الواضح أنني كنت مدركا أن لمعظم الشبكات الإخبارية أجندتها السياسية الخاصة، فإنني أدركت الآن أن لديهم أيضاً تحيزاً متطرفا بشأن جميع القضايا، حتى غير السياسية منها.

أولويات البحث التي عملنا بجد عليها ووضعناها في صورتها النهائية. ومرّرنا تلك الرسالة المفتوحة المؤيدة لأبحاث السلامة (المذكورة في الفصل الأول) على الحضور، وسررنا أن الجميع وقع عليها.

عقدت أنا وميا اجتماعاً سحرياً مع إيلون في غرفتنا بالفندق إذ بارك الخطط التفصيلية لبرنامجنا للمنح البحثية. وتأثرت ميا بمدى تواضعه وصراحته حول حياته الشخصية، ومدى اهتمامه بنا. سألنا كيف التقينا، وأحب قصة ميا التفصيلية. وفي اليوم التالي، صوّرنا معه مقابلة حول سلامة الذكاء الاصطناعي، ولماذا أراد دعمه ماديا، وبدا أن كل شيء يسير على المسار الصحيح.[2]

ذروة المؤتمر، إعلان تبرع إيلون، كان من المقرر أن يكون ذلك في الساعة 7 مساءً، في يوم الأحد، 4 يناير 2015، وكنت متوترا لدرجة أنني تقلّبت في نومي في الليلة السابقة. ثم، وقبل 15 دقيقة فقط قبل الوقت الذي كان من المفترض أن نتوجه إلى الجلسة التي سيعلن فيها عن التبرع واجهتنا عقبة! اتصل مساعد إيلون وقال إنه يبدو أن إيلون ربما لا يكون قادراً على المضي قدماً في الإعلان عن الدعم المالي، وقالت ميا إنها لم ترني قط أكثر إحباطا أو شعورا بخيبة أمل. لكن، أخيرا وصل إيلون، ونحن نجلس هناك ونتحدث سمعت دقات الثواني تبدأ بالعد التنازلي لبدء الجلسة. وأوضح أنهم كانوا على بعد يومين فقط من الإطلاق المهم من مشروع الصاروخ سبيس إكس SpaceX إذ كانوا يأملون أن يطلقوا أول هبوط ناجح على الإطلاق للمرحلة الأولى باستخدام مركبة فضائية من دون طيار، وكان هذا إنجازاً ضخماً، ولم يكن فريق سبيس إكس يرغب في أي ارتباك بفعل تسليط الأضواء الإعلامية على إيلون في تلك الفترة. وأشار أنتوني أغيري Anthony Aguirre، الذي كان لطيفاً ودافئاً كما هو دائماً، أنْ **لا أحد** يريد إثارة اهتمام وسائل الإعلام بهذا الخبر أيضا، لا إيلون ولا مجتمع الذكاء الاصطناعي. وصلنا متأخرين بضع دقائق إلى الجلسة التي كنت أديرها. لكن كانت لدينا خطة: لن يتم ذكر أي مبلغ بالدولار، للتأكد من أن الإعلان لن يثير اهتمام الإعلام، وسأفرض قاعدة تشاتام هاوس على الجميع للحفاظ على سرية إعلان إيلون وإبقائه سرا عن العالم لمدة تسعة أيام إذا وصل صاروخه إلى المحطة الفضائية، بغض النظر عما إذا كان الهبوط ناجحا ؛ وقال إنه سيحتاج إلى مزيد من الوقت إذا انفجر الصاروخ عند الإطلاق.

أخيرا وصل العد التنازلي للإعلان إلى الصفر. وظل المتحدثون في جلسة الذكاء الفائق التي كنت أديرها جالسين واحدا تلو الآخر خشبة المسرح في كراسيهم: اليعازر يودكوفسكي، إيلون موسك، نيك بوستروم، ريتشارد ملاح، موري شاناهان، بارت سلمان Bart Selman، شين ليغ Shane Legg، فيرنور فينغ Vernor Vinge. توقف الناس تدريجيّاً عن التصفيق. لكن أعضاء اللجنة ظلوا جالسين، لأنني طلبت إليهم البقاء دون توضيح السبب. أخبرتني ميا فيما بعد أن صوت ضربات نبضها عند هذه اللحظة وصل إلى الستراترسفير، وأنها قبضت تحت الطاولة على يد فيكتوريا كراكوفنا الباعثة على الهدوء. ابتسمت، لعلمي أن هذه هي اللحظة التي عملنا من أجلها وأملنا بها وانتظرناها.

وخاطبت الجلسة قائلا إنني سعيداً جداً لوجود إجماع في الاجتماع على أن هناك حاجة إلى مزيد من البحث للحفاظ على الذكاء الاصطناعي، وأن هناك الكثير من التوجهات البحثية الملموسة التي يمكننا العمل عليها على الفور. وتابعت قائلا في هذه الجلسة إنه دارت أحاديث عن مخاطر جدية، لذا، سيكون من الجيد رفع معنوياتنا والدخول في مزاج متفائل قبل التوجه إلى مأدبة المؤتمر المقامة في خارج القاعة. «ومن ثم أُعطي الميكروفون لـ...... إيلون ماسك!» شعرت أن التاريخ كان في طور التكوين وإيلون يمسك بالميكرفون ويعلن أنه سيتبرع بمبلغ كبير من المال لأبحاث سلامة الذكاء الاصطناعي. ولم نُدهش من أنه حصل على استحسان القاعة. وكما هو مخطط له، لم يذكر المبلغ، ولكني كنت أعرف أنه كان مبلغاً رائعاً قدره عشرة ملايين دولار، كما اتفقنا.

ذهبت أنا وميا لزيارة والدينا في السويد ورومانيا بعد المؤتمر، وبنفس الأنفاس المحبوسة من التشويق، شاهدنا البث المباشر لإطلاق الصاروخ مع والدي في ستوكهولم. مع الأسف انتهت محاولة الهبوط بما وصفه إيلون ملطّفاً بـ «RUD»، «التفكك السريع غير المجدول» Rapid unscheduled disassembly، واستغرق فريقه 15 شهراً إضافيا لتنفيذ هبوط ناجح للمركبة الفضائية في المحيط.[3] لكن جميع الأقمار الاصطناعية أُطلقت بنجاح لتدور في مداراتها، كما أطلق بنجاح برنامجنا للمنح عبر تغريدة من قبل إيلون للملايين من متابعيه.[4]

تعميم سلامة الذكاء الاصطناعي

أحد الأهداف الرئيسية لمؤتمر بورتوريكو كان إدارج أبحاث سلامة الذكاء الاصطناعي في التيار الرئيسي للأبحاث، وكان من المبهج مشاهدة ذلك يتكشف في عدد من الخطوات. في البداية كان هناك المؤتمر نفسه، حيث أعرب العديد من الباحثين عن ارتياحهم من التعامل مع الموضوع حينما أدركوا أنهم كانوا جزءاً من مجتمع متنامٍ من أقرانهم. وقد تأثرت كثيرا من التشجيع من العديد من المشاركين. مثلاً، أرسل لي أستاذ بارز من جامعة كورنيل، بارت سلمان، رسالة إلكترونية يقول فيها: «بصراحة، لم أر قط اجتماعاً علمياً أفضل تنظيماً أو أكثر إثارة وتحفيزا فكرياً».

بدأت خطوة التعميم التالية في 11 يناير عندما غرد إيلون قائلا: «أفضل مطوري الذكاء الاصطناعي في العالم يوقعون على دعوة مفتوحة لإجراء أبحاث حول سلامة الذكاء الاصطناعي»،[5] مضيفا رابطا صفحة تسجيل سرعان ما جمعت أكثر من ثمانية آلاف توقيع، بما في ذلك العديد من أبرز مطوري الذكاء الاصطناعي في العالم. صار من الصعب بشكل مفاجئ الادعاء بأن الأشخاص المهتمين بسلامة الذكاء الاصطناعي لم يعرفوا ما الذي يتحدثون عنه، لأن هذا يعني ضمناً أن فردا ما من كبار الباحثين في الذكاء الاصطناعي لا يعرف ما الذي يتحدث عنه. وتناقلت وسائل الإعلام في جميع أنحاء العالم الرسالةَ المفتوحة بطريقة جعلتنا نشعر بالامتنان لأننا منعنا الصحفيين من حضور

مؤتمرنا. على الرغم من أن أكثر الكلمات إثارة للانزعاج في الرسالة كانت «المزالق»، فإنها تسببت في عناوين مثل «إيلون ماسك وستيفن هوكينغ يوقعان رسالة مفتوحة على أمل منع انتفاضة الإنسان الآلي»، واستخدموا صورا توضيحية لروبوتات قاتلة. ومن بين مئات من المقالات التي رصدناها، كانت المقالة المفضلة لي مقالة تسخر من المقالات الأخرى، إذ كتبت: «إن عنوانا رئيسيا يستحضر رؤى الروبوتات تشبه الهياكل العظمية التي تدوس بأقدامها جماجم بشرية يحوّل التكنولوجيا المعقدة والتحويلية إلى عرض جانبي في كرنفال».[6] لحسن الحظ كانت هناك العديد من المقالات الإخبارية الرصينة، وضعتنا أمام تحدٍّ آخر: مواكبة سيل من التوقيعات الجديدة، التي تحتاج إلى التحقق يدويا لحماية مصداقيتنا والتخلص من المزح مثل الشخصيات هال "HAL 9000"، "المدمر" Terminator، "سارة جينيت كونور" Sarah Jeanette Connor، و"سكاينت" Skynet. «لهذه الرسالة ولرسائلنا المفتوحة في المستقبل، ساعدت فيكتوريا كراكونفا ويانوس كريمار على تنظيم كتيبة مراجعين من المتطوعين من بينهم جيسي غالف Jesse Galef وإريك غاستريفند Eric Gastfriend ورافاتي فينوث كومار Revathi Vinoth Kumar، يعملون في نوبات، فعندما يخلد رافاتي إلى النوم في الهند، تُمررت عصا القيادة إلى إريك في بوسطن، وهلم جرا.

بدأت خطوة التعميم الثالثة بعد أربعة أيام، عندما غرّد إيلون وأعاد نشر رابطنا الذي نعلن فيه أنه يتبرع بمبلغ 10 ملايين دولار لأبحاث سلامة الذكاء الاصطناعي.[7] بعد أسبوع، أطلقنا بوابة على شبكة الإنترنت يمكن للباحثين من جميع أنحاء العالم التقدم من خلالها والتنافس على هذا التمويل. كنا قادرين على إقامة نظام التقديم بهذه السرعة فقط لأني وأنتوني أمضينا العقد الماضي في إدارة مسابقات مماثلة لمنح الفيزياء. ووافقت مؤسسة مشروع الأعمال الخيرية المفتوحة Open Philanthropy Project، وهي مؤسسة خيرية مقرها كاليفورنيا تركز على المشاريع عميقة التأثير، وافقت بسخط على إضافة تبرع إلى هبة إيلون للسماح لنا بتقديم المزيد من المنح. لم نكن متأكدين من عدد المتقدمين الذين سنحصل عليهم، لأن الموضوع كان جديداً وكان الموعد النهائي قصيراً. غمرنا الرد، إذ تلقينا نحو 300 فريق من جميع أنحاء العالم يطلبون 100 مليون دولار. قامت لجنة من أساتذة الذكاء الاصطناعي والباحثين الآخرين بمراجعة المقترحات بعناية واختارت 37 فريقاً رابحاً تم تمويلهم لمدة تصل إلى ثلاث سنوات. عندما أعلنا قائمة الفائزين، كان ذلك بمثابة المرة الأولى التي كانت فيها استجابة وسائل الإعلام لأنشطتنا دقيقة إلى حد ما وخالية من صور الروبوت القاتل. فقد أدركوا أخيرا أن سلامة الذكاء الاصطناعي لم تكن حديثاً فارغاً: كان هناك عمل مفيد فعلي يتعين القيام به، وكان الكثير من فرق البحث الكبرى تُشمّر عن سواعدها للانضمام إلى هذا الجهد.

واستكملت خطوة التعميم الرابعة عبر العامين المقبلين، بنشر عشرات من المنشورات التقنية وإقامة عشرات من ورش العمل حول سلامة الذكاء الاصطناعي في جميع

أنحاء العالم، وعادة ما تكون ضمن فعاليات مؤتمرات الذكاء الاصطناعي الرئيسية. لقد حاول الأشخاص الدؤوبون لسنوات عديدة إشراك مجتمع الذكاء الاصطناعي في أبحاث السلامة، ولم يحققوا سوى نجاح محدود لكن الأمور بدأت تأخذ مجراها فعلاً الآن. موّل برامجنا للمنح العديد من هذه المنشورات، وقد بذلنا في المعهد FLI قصارى جهدنا للمساعدة على تنظيم وتمويل أكبر عدد ممكن من ورش العمل هذه. لكن تحققت مجموعة متزايدة من هذه الأبحاث بفعل تخصيص باحثي الذكاء الاصطناعي وقتهم ومواردهم للقيام بذلك. ونتيجة لذلك، تعلم المزيد من الباحثين حول أبحاث السلامة من زملائهم، واكتشفوا أنه بصرف النظر عن كونها مفيدة، يمكن قد تكون ممتعة أيضاً، إذ تنطوي على مشكلات رياضياتية وحوسبية مثيرة للاهتمام.

لا يوافق الجميع على أن العمل مع المعادلات المعقدة عمل مرح. بعد عامين من مؤتمر بورتوريكو، أقمنا ورشة عمل تقنية قبل مؤتمرنا الآسيوي عرض فيه الفائزون بمنح المعهد FLI أبحاثهم، وشاهدنا الشريحة بعد الشريحة من العروض التقديمية المفعمة بالرموز الرياضياتية على الشاشة الكبيرة. موشيه فاردي Moshe Vardi، وهو أستاذ الذكاء الاصطناعي في جامعة رايس Rice University، تندر قائلا إنه علم أننا نجحنا في تأسيس أبحاث حول سلامة الذكاء الاصطناعي بمجرد أن صارت الاجتماعات مملة.

لم يقتصر هذا النمو الهائل في أبحاث سلامة الذكاء الاصطناعي على الأكاديميين. بل أطلقت أمازون وديب مايند و فيسبوك وغوغل وأي بي أم و ميكروسوفت شراكة صناعية للذكاء الاصطناعي المفيد.[8] وقد مكنت تبرعات جديدة رئيسية الأبحاثَ الموسعة في أكبر المنظمات الشقيقة غير التابعة لنا: معهد أبحاث الذكاء الآلي في بيركلي، ومعهد مستقبل الإنسانية في أكسفورد، ومركز دراسة المخاطر الوجودية في كامبريدج (المملكة المتحدة)، والمزيد من التبرعات بقيمة عشرة ملايين دولار أو أكثر من الجهود الإضافية المفيدة: مركز ليفيرهولم لمستقبل الذكاء Leverhulme Centre for the Future of Intelligence في كامبريدج، و وقفية كي. و إل. غيتس للأخلاق والتقنيات الحوسبية K&L Gates Endowment for Ethics and Computational Technologies في بيتسبيرغ، وصندوق أخلاق وحوكمة الذكاء الاصطناعي Ethics and Governance of Artificial Intelligence Fund في ميامي. أخيراً وليس آخراً، مع التزام ببليون دولار، تعاون إيلون موسك مع رواد أعمال آخرين لإطلاق شركة OpenAI، وهي شركة غير ربحية في سان فرانسيسكو، تسعى إلى تطوير الذكاء الاصطناعي المفيد، ترسخت أبحاث السلامة الذكاء الاصطناعي.

بالتزامن مع هذه الطفرة في البحث، ظهرت موجة من الآراء، سواء الفردية أم الجماعية. ونشرت مجموعة الصناعة حول الذكاء الاصطناعي Partnership on AI مبادئها التأسيسية، ونشرت حكومة الولايات المتحدة وجامعة ستانفورد والمنظمة IEEE (وهي أكبر منظمة في العالم للمهنيين التقنيين) تقاريرها الطويلة، إضافة إلى عشرات التقارير وورقات المواقف الأخرى Position papers، في العديد من الأماكن الأخرى.[9]

في أسيلومار كنا حريصين على إدارة مناقشة هادفة بين الحاضرين، ومعرفة ما يتفق عليه هذا المجتمع المتنوع، إن وجد مثل هذا الاتفاق. ومن ثم، تولى لوكاس بيري المهمة البطولية المتمثلة في قراءة جميع تلك الوثائق واستخلاص الآراء. وفي جهد ماراثوني بدأه أنتوني أغيري، واختتم بسلسلة من الاجتماعات الهاتفية الطويلة، حاول فريق المعهد FLI لدينا بعد ذلك تجميع الآراء المتشابهة معاً والتخلص من الإسهابات البيروقراطية الزائدة عن الحاجة، إلى أن انتهى الفريق إلى قائمة واحدة من المبادئ المختصرة، بما في ذلك أيضاً آراء مؤثرة ولكن غير منشورة وتم التعبير عنها بشكل غير رسمي في المحادثات وما إلى ذلك. غير أن هذه اللائحة شملت الكثير من الغموض والتناقض وما هو عرضة للتفسير، لذا في الشهر الذي سبق المؤتمر، أرسلنا اللائحة إلى جميع المشاركين وجمعنا آراءهم واقتراحاتهم لتحسين المبادئ أو لمبادئ جديدة. أنتجت هذه المساهمة المجتمعية قائمة مبدئية منقحة بشكل كبير لاستخدامها في المؤتمر.

وفي أسيلومار، تم تنقيح القائمة في خطوتين. أولاً، ناقشت مجموعات صغيرة المبادئ التي كانت أكثر اهتماماً بها (الشكل 4.9)، منتجين توصيات مفصلة، وآراء، ومبادئ جديدة ونسخا متنافسة من التوصيات القديمة. أخيراً، استطلعنا رأي جميع الحاضرين لتحديد مستوى الدعم لكل نسخة من كل مبدأ.



الشكل 3.9: في أسيلومار مجموعات من العقول العظيمة تتأمل مبادئ الذكاء الاصطناعي.

كانت هذه العملية الجماعية شاملة ومرهقة على حد سواء، إذ قمت أنا وأنتوني وميا بتقليص وقت النوم والغداء في المؤتمر حرصا على تجميع كل ما هو مطلوب في الوقت المناسب للخطوات التالية. لكنها كانت فترة مثيرة أيضا، فبعد هذه المناقشات التفصيلية والشائكة والمثيرة للجدل في بعض الأحيان ومثل هذه المجموعة الواسعة

من الآراء، أدهشنا المستوى العالي من الإجماع الذي ظهر حول العديد من المبادئ خلال تلك الدراسة الاستقصائية، مع حصول البعض منها على دعم أكثر من 97%. وسمح لنا هذا الإجماع بوضع شروط عالية للإدراج في القائمة النهائية: لقد احتفظنا بالمبادئ التي وافق عليها 90% على الأقل من الحاضرين، على الرغم من أن هذا يعني أن بعض المبادئ العامة تم إسقاطها في اللحظة الأخيرة، بما في ذلك بعض المبادئ المفضلة لدي أنا شخصيا،[10] وهذا جعل معظم المشاركين يشعرون بالارتياح من تأييدها جميعاً والتوقيع على ورقة الموافقات التي مررناها حول قاعة محاضرات، ها هي النتيجة.

مبادئ أسيلومار للذكاء الاصطناعي

وفر الذكاء الفني بالفعل أدوات مفيدة يتم استخدامها يومياً من قبل أشخاص من جميع أنحاء العالم، وسيوفر تطويرها المستمر، مسترشداً بالمبادئ التالية، فرصاً رائعة لمساعدة وتمكين الناس في العقود والقرون القادمة.

قضايا البحث

§ 1 هدف البحث: يجب أن يكون الهدف من بحث الذكاء الاصطناعي هو خلق ذكاء غير محدود، ولكن يكون ذكاء مفيدا.

§ 2 تمويل الأبحاث: يجب أن تكون الاستثمارات في الذكاء الاصطناعي مصحوبة بتمويل للبحث عن ضمان الاستخدام المفيد له، بما في ذلك معالجة الأسئلة الشائكة في علوم الحاسوب والاقتصاد والقانون والأخلاقيات والدراسات الاجتماعية، مثل:

(أ) كيف يمكننا أن نجعل أنظمة الذكاء الاصطناعي في المستقبل منيعة، بحيث تفعل ما نريد دون حدوث خلل أو اختراق؟

(ب)كيف يمكننا تنمية ازدهارنا من خلال الأتمتة مع الحفاظ على موارد الناس وهدفهم؟

(ج) كيف يمكننا تحديث أنظمتنا القانونية لتكون أكثر عدالة وفعالية، لمواكبة الذكاء الاصطناعي، ولإدارة المخاطر المرتبطة بالذكاء الاصطناعي؟

(د) ما هي مجموعة القيم التي يجب أن يتماشى معها الذكاء الاصطناعي، وما هو الوضع القانوني والأخلاقي الذي ينبغي أن يكون عليه؟

§ 3 رابطة السياسة-العلوم: يجب أن يكون هناك تبادل بنّاء وصحي بين الباحثين في الذكاء الاصطناعي وواضعي السياسات.

§ 4 ثقافة البحث: يجب تعزيز ثقافة التعاون والثقة والشفافية بين الباحثين ومطوري الذكاء الاصطناعي.

§ 5 تجنب السباق: يجب أن تتعاون الفرق التي تطور أنظمة الذكاء الاصطناعي بنشاط لضمان عدم اختزال المسافات فيما يختص بمعايير السلامة.

أخلاق وقيم

§ 6 السلامة: يجب أن تكون أنظمة الذكاء الاصطناعي آمنة ومأمونة طوال فترة تشغيلها، وينبغي التحقق مما إذا كان ذلك قابلاً للتطبيق ومجدياً.

§ 7 شفافية الفشل: إذا تسبب نظام الذكاء الاصطناعي في حدوث ضرر، فيجب أن يكون من الممكن التأكد من السبب.

§ 8 الشفافية القضائية: أي تدخل من نظام ذاتي التحكم في صنع القرار القضائي يجب أن يقدم تفسيرا مُرضيا قابلا للمراجعة من قبل سلطة بشرية مختصة.

§ 9 المسؤولية: المصممون وبناة أنظمة الذكاء الاصطناعي المتطورة هم أصحاب مصلحة في الآثار الأخلاقية المترتبة على استخدامها وإساءة استخدامها وأفعالها، مع مسؤولية، ولديهم فرصة لتشكيل تلك التأثيرات وما تتضمنه من آثار.

§ 10 مواءمة القيمة: يجب تصميم أنظمة الذكاء الاصطناعي عالية الاستقلالية بحيث يمكن ضمان أهدافها وسلوكياتها للتوافق مع القيم الإنسانية عند تشغيلها.

§ 11 القيم الإنسانية: يجب تصميم أنظمة الذكاء الاصطناعي وتشغيلها بحيث تتوافق مع مُثُل كرامة الإنسان وحقوقه وحرياته وتنوعه الثقافي.

§ 12 الخصوصية الشخصية: يجب أن يتمتع الأشخاص بالحق في الوصول إلى البيانات التي يقومون بإنشائها وإدارتها والتحكم فيها، مع إعطاء قدرة أنظمة الذكاء الاصطناعي على تحليل تلك البيانات واستخدامها.

§ 13 الحرية والخصوصية: يجب ألا يؤدي تطبيق ذكاء اصطناعي على البيانات الشخصية إلى الحد بشكل غير معقول من حرية الأشخاص الحقيقية أو المتصورة.

§ 14 الفوائد المشتركة: يجب أن تُفيد تقنيات الذكاء الاصطناعي وتمكّن أكبر عدد ممكن من الأشخاص.

§ 15 الازدهار المشترك: ينبغي تقاسم الرخاء الاقتصادي الذي يُنشئه ذكاء اصطناعي على نطاق واسع، لفائدة البشرية جمعاء.

§ 16 التحكم البشري: يجب أن يختار البشر كيفية وما إذا كان سيتم تفويض اتخاذ القرارات بأنظمة الذكاء الاصطناعي، لتحقيق الأهداف التي اختارها الإنسان.

§17 عدم التقويض: يجب على السلطة الناجمة من تحكم أنظمة ذكاء اصطناعي متطورة احترام وتحسين، بدلاً من تقويض، العمليات الاجتماعية والمدنية التي تعتمد عليها صحة المجتمع.

§ 18 سباق تسلح ذكاء اصطناعي: يجب تجنب سباق التسلح في الأسلحة الذاتية الفتاكة.

قضايا طويلة الأجل

§ 19 تحذير القدرة: لما لم يتحقق إجماع، يجب أن نتجنب الافتراضات فيما يتعلق بالحدود العليا لقدرات الذكاء الاصطناعي في المستقبل.

§ 20 الأهمية: يمكن أن يُحدث الذكاء الاصطناعي المتطور تغييرا عميقا في تاريخ الحياة على الأرض، وينبغي التخطيط له وإدارته بتوفير العناية والموارد اللازمة.

§21 المخاطر: يجب أن تخضع المخاطر المرتبطة بأنظمة الذكاء الاصطناعي، وخاصة المخاطر الكارثية أو الوجودية، لجهود التخطيط وتخفيف الضرر Mitigation بما يتناسب مع تأثيرها المتوقع.

§ 22 التحسين الذاتي المتكرر: أنظمة ذكاء اصطناعي مصممة لتحسين نفسها أو نسخ نفسها نسخا طبق الأصل ذاتياً بشكل متكرر وبطريقة قد تؤدي بسرعة إلى زيادة الجودة أو الكمية التي يجب أن تخضع لقواعد سلامة وتحكم صارمة.

§ 23 الصالح العام: يجب تطوير الذكاء الفائق فقط لخدمة المثل الأخلاقية المشتركة على نطاق واسع، ولصالح البشرية جمعاء بدلاً من دولة واحدة أو منظمة واحدة.

نمت قائمة التوقيعات بشكل كبير بعد نشرنا المبادئ على الإنترنت، وهي تضم الآن قائمة مذهلة تضم أكثر من ألف من باحثي الذكاء الاصطناعي والعديد من كبار المفكرين الآخرين، إذا كنت ترغب أيضاً في الانضمام كموقِّع، فيمكنك القيام بذلك على الموقع: http://futureoflife.org/ai-principles.

لقد أدهشنا ليس فقط مستوى الإجماع حول المبادئ، ولكن أيضاً مدى قوتها. بالتأكيد، يبدو بعضها من أول نظرة مثيراً للجدل مثل «السلام والحب والأمومة هي أمور قيّمة». لكن العديد من هذه المبادئ لها أنياب حقيقية [تفرض تحقيقها فرضا]، كما يُرى بسهولة أكبر عند صياغة نفي لها. مثلاً، «الذكاء الفائق أمر مستحيل!» ينتهك الفقرة 19، و«إنها مضيعة تامة لإجراء الأبحاث على الحد من المخاطر الوجودية للذكاء الاصطناعي!» تنتهك الفقرة 21.

بالطبع، كما يمكنك أن ترى بنفسك إذا كنت شاهدت حلقة نقاشنا طويلة الأجل على يوتيوب،[11] إيلون ماسك، وستيوارت رسل، وسام هاريس، ونيك بوستروم، وديفيد تشالمرز، وبارت سيلمان، ويان تالين، جميعهم اتفقوا على أنه من المحتمل جدّاً تطوير ذكاء فائق وأن أبحاث السلامة مهمة.

تفاؤل واعي

كما اعترفت في افتتاح هذا الخاتمة، أشعر أنني أكثر تفاؤلا بمستقبل الحياة مما كنت. دعوني أشاطركم قصة شخصية توضح سبب ذلك.

زادت خبراتي خلال السنوات القليلة الماضية من تفاؤلي لسببين منفصلين. أولاً، شهدت تعاون مجتمع الذكاء الاصطناعي تعاونا رائعا للمضي قدماً في مجابهة التحديات المقبلة، وفي كثير من الأحيان بالتعاون مع مفكرين من حقول أخرى. أخبرني إيلون بعد اجتماع أسيلومار أنه وجد أنه من المذهل كيف انتقلت سلامة الذكاء الاصطناعي من قضية هامشية لتصبح قضية تهم التيار الرئيسي في غضون بضع سنوات فقط، وأنا نفسي مدهوش بالقدر نفسه. والآن، ليس الأمر مجرد قضايا المدى القريب من الفصل

الثالث هي التي صارت موضوعات نقاش ذات أهمية، بل صار حتى الذكاء الفائق والمخاطر الوجودية، كما هي الحال في مبادئ أسيلومار للذكاء الاصطناعي. ولم يكن بالإمكان بأي طريقة تبني تلك المبادئ في بورتوريكو قبل عامين، إذ شقت كلمة "المزالق" Pitfalls الأكثر رعبا طريقها لتدخل في نص الرسالة المفتوحة.

أحب مشاهدة الناس، وفي مرحلة ما خلال صباح اليوم الأول من مؤتمر أسيلومار، وقفت إلى جانب القاعة وشاهدت المشاركين يستمعون إلى مناقشة حول الذكاء الاصطناعي والقانون. لدهشتي، اجتاحني شعور دافئ وغامض وشعرت أنني متحمس للهدف. كان الشعور مختلفا تماماً عن بورتوريكو! في ذلك الوقت، أتذكر أنني نظرت إلى معظم مجتمع الذكاء الاصطناعي بمزيج من الاحترام والخوف - ليس فقط كفريق معارض، ولكن كمجموعة شعرت أنا وزملائي المعنيين بالذكاء الاصطناعي بأننا بحاجة إلى إقناعهم. لكن الآن بدا الأمر واضحاً أننا كنا جميعاً في الفريق **نفسه**. وربما كما استخلصت من قراءة هذا الكتاب، ما زلت لا أمتلك إجابات لكيفية إنشاء مستقبل رائع مع الذكاء الاصطناعي، لذلك فهو شعور رائع أن تكون جزءاً من مجتمع متنام يبحث عن الإجابات معاً.



الشكل 4.9: في أسيلومار معاً يبحث المجتمع المتنامي عن الإجابات.

السبب الثاني الذي يجعلني أكثر تفاؤلاً هو أن تجربة المعهد FLI كانت تجربة مُمكّنة. إن ما أثار دموعي في لندن كان شعوراً بالحتمية: قد يكون هناك مستقبل مزعج قادم ولم يكن هناك شيء يمكننا القيام به حيال ذلك. لكن السنوات الثلاث التالية أذابت كآبتي القاتمة. إذا كان بإمكان شرذمة من المتطوعين غير مدفوعي الأجر أن تحدث فرقاً إيجابياً فيما يمكن القول إنه أهم محادثة في عصرنا، تخيل ما يمكن أن نفعله جميعاً إذا عملنا معاً!

تحدث إريك برينجولفسون عن نوعين من التفاؤل في حديثه عن أسيلومار: أولاً، هناك هذا النوع غير المشروط، مثل التوقع الإيجابي بأن الشمس ستشرق صباح غد، ثم هناك ما أسماه "التفاؤل الواعي" Mindful optimism، وهو توقع حدوث الأشياء الجيدة إذا خططت بعناية وعملت بجد من أجل تحقيقها، هذا هو نوع التفاؤل الذي أشعر به الآن بشأن مستقبل الحياة.

إذاً ما الذي **يمكنك** فعله لإحداث تغيير إيجابي لمستقبل الحياة مع دخولنا عصر الذكاء الاصطناعي؟ لأسباب سأشرحها قريباً، أعتقد أن الخطوة الأولى هي العمل على أن تصبح متفائلاً متيقناً، إذا لم تكن واحداً بالفعل. فلكي تكون متفائلاً ناجحاً في التفكير، من الضروري تطوير رؤى إيجابية للمستقبل. عندما يأتي طلبة المعهد MIT إلى مكتبي بحثا عن المشورة المهنية، عادة ما أبدأ بسؤالهم: أين يرون أنفسهم خلال عقد من الزمن، إذا أجاب أحد الطلبة: «ربما سأكون في جناح للسرطان، أو في مقبرة بعد اصطدامي بحافلة». فإنني أصعّب الأمور عليه. فتصور عقود قادمة سلبية فقط هو مقاربة فظيعة للتخطيط الوظيفي! إن تكريس 100% من جهود المرء لتجنب الأمراض والحوادث هو وصفة عظيمة للوسواس المرضي Hypochondria والارتياب Paranoia، وليس السعادة. بدلاً من ذلك، أود أن أسمع الطالب يصف أهدافه بحماس، وبعد ذلك يمكننا مناقشة استراتيجيات الوصول إلى هناك مع تجنب المزالق.

أشار إيريك إلى أنه وفقا لنظرية الألعاب Game theory لماذا التضحية بشيء لديك إذا كنت لا تستطيع أن تتخيل كسبا أكبر سيحقق من ذلك؟ هذا يعني أننا يجب أن نتخيل مستقبلاً إيجابياً ليس فقط لأنفسنا، ولكن أيضاً للمجتمع الإنساني نفسه. بعبارة أخرى، نحن بحاجة إلى المزيد من الأمل الوجودي. كما تحب ميا أن تذكرني، من فرانكشتاين إلى المُدمِّر، فإن الرؤى المستقبلية في الأدب والسينما هي في الغالب لاطوباوية Dystopic. وبعبارة أخرى، نحن كمجتمع نخطط لمستقبلنا بنفس رداءة تخطيط هذا الطالب الافتراضي من طلبة المعهد MIT. لذا نحن بحاجة إلى مزيد من التفاؤل الواعي. و لهذا السبب شجعتك في هذا الكتاب على التفكير في أي نوع من المستقبل **تريد** وليس مجرد أي نوع من المستقبل أنت **خائف** منه، بحيث تجد أهدافاً مشتركة للتخطيط والعمل من أجل تحقيقها.

عبر هذا الكتاب، رأينا كيف من المحتمل أن يمنحنا الذكاء الاصطناعي فرصا كبرى وتحديات صعبة على حد سواء. والاستراتيجية التي قد تساعد على حل جميع تحديات الذكاء الاصطناعي هي أن نُنظّم أنفسنا ونطور مجتمعنا الإنساني **قبل** أن ينطلق الذكاء الاصطناعي انطلاقة كاملة. يتعين علينا تثقيف شبابنا لجعل التكنولوجيا متينة ومفيدة قبل التخلي لها عن قوة كبيرة. ويتعين علينا تحديث قوانيننا قبل أن تجعلها التكنولوجيا قديمة. ويتعين علينا حل النزاعات الدولية قبل أن تتحول إلى سباق تسلح لأسلحة ذاتية التحكم. ويتعين علينا أن نخلق اقتصاداً يضمن الرخاء للجميع قبل أن تتفاقم أوجه عدم المساواة المحتملة. سنكون في وضع أفضل في مجتمع يتم فيه تنفيذ نتائج أبحاث سلامة الذكاء الاصطناعي بدلاً من تجاهلها. وعند التطلع إلى الأمام، وللتحديات المتعلقة

بالذكاء AGI الفائق، من الأفضل أن نتفق على بعض المعايير الأخلاقية الأساسية على الأقل قبل أن نبدأ في تدريس هذه المعايير لآلات قوية. ففي عالم مُستقطَب وفوضوي، سيكون لدى الأشخاص الذين لديهم القدرة على استخدام الذكاء الاصطناعي لأغراض خبيثة قدر أكبر من الحافز والقدرة على القيام بذلك، وستشعر الفرق التي تتسابق لبناء الذكاء AGI بمزيد من الضغط لاختصار المراحل وعدم الالتفات إلى محاذير السلامة بدلاً من التعاون. باختصار، إذا استطعنا إنشاء مجتمع إنساني أكثر تناسقاً يتمسك بالتعاون من أجل تحقيق الأهداف المشتركة، فسيؤدي ذلك إلى تحسين احتمالات أن تنتهي ثورة الذكاء الاصطناعي إلى ما هو مفيد.

بعبارة أخرى، أحد أفضل الطرق لتحسين مستقبل الحياة هو أن تحسّن الغد. لديك القدرة على فعل ذلك بعدة طرق. بالطبع يمكنك التصويت وإخبار الساسة عن رأيك في التعليم والخصوصية والأسلحة الفتاكة ذاتية التحكم والبطالة التكنولوجية وغيرها من القضايا. لكنك أيضاً تصوّت يومياً من خلال ما تختاره للشراء، والأخبار التي تختار أن تستهلكها، وما تختار مشاركته، ونوع نمط الدور الذي تختاره، هل تريد أن تكون شخصاً يقاطع جميع محادثاته للتحقق من هاتفه الذكي أو شخصا يشعر بالتمكين لاستخدامه التكنولوجيا بطريقة مُخطّط لها ومقصودة؟ هل ترغب في امتلاكك التكنولوجيا الخاصة بك أم هل ترغب في أن تمتلكك التكنولوجيا؟ ماذا تريد أن يعني أن تكون إنساناً في عصر الذكاء الاصطناعي؟ يرجى مناقشة كل هذا مع من حولك - إنها ليست فقط محادثة مهمة، ولكنها محادثة رائعة.

نحن حماة مستقبل الحياة الآن ونحن نصوغ عصر الذكاء الاصطناعي. وعلى الرغم من أنني بكيت في لندن، أشعر الآن أنه لا يوجد شيء لا مفر منه في المستقبل، وأنا أعلم أن إحداث فرق هو أسهل بكثير مما كنت أعتقد. مستقبلنا ليس منقوشا في الحجر وبانتظار أن يحدث لنا - إنه مستقبلنا، دعونا نخلق مستقبلا ملهما معا!

الهوامش

الفصل الأول

1. “The AI Revolution: Our Immortality or Extinction?” *Wait But Why* (January27, 2015), at http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2 .html.
2. This open letter, “Research Priorities for Robust and Beneficial Artificial Intelligence,” can be found at http://futureoflife.org/ai-open-letter/.
3. Example of classic robot alarmism in the media: Ellie Zolfagharifard, “Artificial Intelligence ‘Could Be the Worst Thing to Happen to Humanity,’ ” *Daily Mail*, May 2, 2014; http://tinyurl.com/hawkingbots.

الفصل الثاني

1. Notes on the origin of the term AGI: http://wp.goertzel.org/who-coined-the -term-agi.
2. Hans Moravec, “When Will Computer Hardware Match the Human Brain?” *Journal of Evolution and Technology* (1998), vol. 1.
3. In the figure showing computing power versus year, the pre-2011 data is from Ray Kurzweil’s book *How to Create a Mind*, and subsequent data is computed from the references in https://en.wikipedia.org/wiki/FLOPS.
4. Quantum computing pioneer David Deutsch describes how he views quantum computation as evidence of parallel universes in his *The Fabric of Reality: The Science of Parallel Universes—and Its Implications* (London: Allen Lane, 1997). If you want my own take on quantum parallel universes as the third of four multiverse levels, you’ll find it in my previous book: Max Tegmark, Our Mathematical Universe: *My Quest for the Ultimate Nature of Reality* (New York: Knopf, 2014).

الفصل الثالث

1. Watch “Google DeepMind’s Deep Q-learning Playing Atari Breakout” on YouTube at https://tinyurl.com/atariai.
2. See Volodymyr Mnih et al., “Human-Level Control Through Deep Reinforcement Learning,” *Nature* 518 (February 26, 2015): 529–533, available online at http://tinyurl.com/ataripaper.
3. Here's a video of the Big Dog robot in action: https://www.youtube.com/ watch?v=W1czBcnX1Ww.
4. For reactions to the sensationally creative line 5 move by AlphaGO, see “Move 37!! Lee Sedol vs AlphaGo Match 2,” at https://www.youtube.com/ watch?v=JNrXgpSEEIE.
5. Demis Hassabis describing reactions to AlphaGo from human Go players: https://www.youtube.com/watch?v=otJKzpNWZT4.
6. For recent improvements in machine translation, see Gideon Lewis-Kraus, “The Great A.I. Awakening,” *New York Times Magazine* (December 14, 2016), available online at http://www.nytimes.com/2016/12/14/magazine/the-great-ai -awakening.html. GoogleTranslate is available here at https://translate.google.com.

7. Winograd Schema Challenge competition: http://tinyurl.com/winogradchal lenge.
8. Ariane 5 explosion video: https://www.youtube.com/watch?v=qnHn8W1Em6E.
9. Ariane 5 Flight 501 Failure report by the inquiry board: http://tinyurl.com/ arianeflop.
10. NASA's Mars Climate Orbiter Mishap Investigation Board Phase I report: http://tinyurl.com/marsflop.
11. The most detailed and consistent account of what caused the Mariner 1 Venus mission failure was incorrect hand-transcription of a single mathematical symbol (a missing overbar): http://tinyurl.com/marinerflop.
12. A detailed description of the failure of the Soviet Phobos 1 Mars mission can be found in Wesley T. Huntress Jr. and Mikhail Ya. Marov, *Soviet Robots in the Solar System* (New York: Praxis Publishing, 2011), p. 308.
13. How unverified software cost Knight Capital $440 million in 45 minutes: http:// tinyurl.com/ knightflop1 and http://tinyurl.com/knightflop2.
14. U.S. government report on the Wall Street "flash crash": "Findings Regarding the Market Events of May 6, 2010" (September 30, 2010), at http://tinyurl.com/ flashcrashreport.
15. 3-D printing of buildings (https://www.youtube.com/watch?v=SObzNdyRTBs), micromechanical devices (http://tinyurl.com/tinyprinter) and many things in between (https://www.youtube.com/watch?v=xVU4FLrsPXs).
16. Global map of community-based fab labs: https://www.fablabs.io/labs/map.
17. News article about Robert Williams being killed by an industrial robot: http:// tinyurl.com/ williamsaccident.
18. News article about Kenji Urada being killed by an industrial robot: http:// tinyurl.com/ uradaaccident.
19. News article about Volkswagen worker being killed by an industrial robot: http://tinyurl.com/baunatalaccident.
20. U.S. government report on worker fatalities: https://www.osha.gov/dep/fatcat/ dep_fatcat.html.
21. Car accident fatality statistics: http://tinyurl.com/roadsafety2 and http://tinyurl .com/roadsafety3.
22. On the first Tesla autopilot fatality, see Andrew Buncombe, "Tesla Crash: Driver Who Died While on Autopilot Mode 'Was Watching Harry Potter,'" *Independent* (July 1, 2016), http:// tinyurl.com/teslacrashstory. For the report of the Office of Defects Investigation of the U.S. National Highway Traffic Safety Administration, see http://tinyurl.com/teslacrashreport.
23. For more about the *Herald of Free Enterprise* disaster, see R. B. Whittingham, The Blame Machine: Why Human Error Causes Accidents (Oxford, UK: Elsevier, 2004).
24. Documentary about the Air France 447 crash: https://www.youtube.com/ watch?v=dpPkp8OGQFI; accident report: http://tinyurl.com/af447report; outside analysis: http://tinyurl.com/ thomsonarticle.
25. Official report on the 2003 U.S.-Canadian blackout: http://tinyurl.com/ uscanadablackout.
26. Final report of the President's Commission on the Accident at Three Mile Island: http:// www.threemileisland.org/downloads/188.pdf.
27. Dutch study showing how AI can rival human radiologists at MRI-based diagnosis of prostate cancer: http://tinyurl.com/prostate-ai.
28. Stanford study showing how AI can best human pathologists at lung cancer diagnosis: http://tinyurl.com/lungcancer-ai.
29. Investigation of the Therac-25 radiation therapy accidents: http://tinyurl.com/ theracfailure.
30. Report on lethal radiation overdoses in Panama caused by confusing user interface: http:// tinyurl.com/cobalt60accident.
31. Study of adverse events in robotic surgery: https://arxiv.org/abs/1507.03518.
32. Article on number of deaths from bad hospital care: http://tinyurl.com/medac cidents.

33. Yahoo set a new standard for "big hack" when announcing a billion(!) of their user accounts had been breached: https://www.wired.com/2016/12/yahoo-hack -billion-users/.
34. *New York Times* article on acquittal and later conviction of KKK murderer: http://tinyurl. com/kkkacquittal.
35. The Danziger et al. 2011 study (http://www.pnas.org/content/108/17/6889 .full), arguing that hungry judges are harsher, was criticized as flawed by Keren Weinshall-Margela and John Shapard (http ://www.pnas.org/content/108/42/ E833.full), but Danziger et al. insist that their claims remain valid (http://www .pnas.org/content/108/42/E834.full).
36. *Pro Publica* report on racial bias in recidivism-prediction software: http://tinyurl .com/ robojudge.
37. Use of fMRI and other brain-scanning techniques as evidence in trials is highly controversial, as is the reliability of such techniques, although many teams claim accuracies better than 90%: http://journal.frontiersin.org/article/10.3389/fpsyg .2015.00709/full.
38. PBS made the movie *The Man Who Saved the World* about the incident where Vasili Arkhipov single-handedly prevented a Soviet nuclear strike: https://www .youtube.com/ watch?v=4VPY2SgyG5w.
39. The story of how Stanislav Petrov dismissed warnings of a U.S. nuclear attack as a false alarm was turned into the movie *The Man Who Saved the World* (not to be confused with the movie by the same title in the previous note), and Petrov was honored at the United Nations and given the World Citizen Award: https:// www.youtube.com/watch?v=IncSjwWQHMo.
40. Open letter from AI and robotics researchers about autonomous weapons: http://futureoflife. org/open-letter-autonomous-weapons/.
41. A U.S. official seemingly wanting a military AI arms race: http://tinyurl.com/ workquote.
42. Study of wealth inequality in the United States since 1913: http://gabriel -zucman.eu/files/ SaezZucman2015.pdf.
43. Oxfam report on global wealth inequality: http://tinyurl.com/oxfam2017.
44. For a great introduction to the hypothesis of technology-driven inequality, see Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (New York: Norton, 2014).
45. Article in *The Atlantic* about falling wages for the less educated: http://tinyurl .com/wagedrop.
46. The data plotted are taken from Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez and Gabriel Zucman, *The World Wealth and Income Database* (http:// www.wid.world), including capital gains.
47. Presentation by James Manyika showing income shifting from labor to capital: http:// futureoflife.org/data/PDF/james_manyika.pdf.
48. Forecasts about future job automation from Oxford University (http:// tinyurl.com/ automationoxford) and McKinsey (http://tinyurl.com/automation mckinsey).
49. Video of robotic chef: https://www.youtube.com/watch?v=fE6i2OO6Y6s.
50. Marin Soljabiü explored these options at the 2016 workshop Computers Gone Wild: Impact and Implications of Developments in Artificial Intelligence on Society: http://futureoflife. org/2016/05/06/computers-gone-wild/.
51. Andrew McAfee's suggestions for how to create more good jobs: http://future oflife.org/ data/PDF/andrew_mcafee.pdf. 52. In addition to many academic articles arguing that "this time is different" for technological unemployment, the video "Humans Need Not Apply" succinctly makes the same point: https://www.youtube.com/watch?v=7Pq-S557XQU.
53. U.S. Bureau of Labor Statistics: http://www.bls.gov/cps/cpsaat11.htm.
54. Argument that "this time is different" for technological unemployment: Federico Pistono, *Robots Will Steal Your Job, but That's* OK (2012), http://robotswill stealyourjob.com.

55. Changes in the U.S. horse population: http://tinyurl.com/horsedecline.
56. Meta-analysis showing how unemployment affects well-being: Maike Luhmann et al., "Subjective Well-Being and Adaptation to Life Events: A Meta-Analysis," *Journal of Personality and Social Psychology* 102, no. 3 (2012): 592; available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289759.
57. Studies of what boosts people's sense of well-being: Angela Duckworth, Tracy Steen and Martin Seligman, "Positive Psychology in Clinical Practice," *Annual Review of Clinical Psychology* 1 (2005): 629–651, online at http://tinyurl.com/well beingduckworth. Weiting Ng and Ed Diener, "What Matters to the Rich and the Poor? Subjective Well-Being, Financial Satisfaction, and Postmaterialist Needs Across the World," *Journal of Personality and Social Psychology* 107, no. 2 (2014): 326, online at http://psycnet.apa.org/journals/psp/107/2/326. Kirsten Weir, "More than Job Satisfaction," *Monitor on Psychology* 44, no. 11 (December 2013), online at http://www.apa.org/monitor/2013/12/job-satisfaction.aspx.
58. Multiplying together about 10^{11} neurons, about 10^4 connections per neuron and about one (10^0) firing per neuron each second might suggest that about 10^{15} FLOPS (1 petaFLOPS) suffice to simulate a human brain, but there are many poorly understood complications, including the detailed timing of firings and the question of whether small parts of neurons and synapses need to be simulated too. IBM computer scientist Dharmendra Modha has estimated that 38 petaFLOPS are required (http://tinyurl.com/javln43), while neuroscientist Henry Markram has estimated that one needs about 1,000 petaFLOPS (http:// tinyurl.com/6rpohqv). AI researchers Katja Grace and Paul Christiano have argued that the most costly aspect of brain simulation is not computation but *communication*, and that this too is a task in the ballpark of what the best current supercomputers can do: http://aiimpacts.org/about.
59. For an interesting estimate of the computational power of the human brain: Hans Moravec "When Will Computer Hardware Match the Human Brain?" *Journal of Evolution and Technology*, vol. 1 (1998).

الفصل الرابع

1. For a video of the first mechanical bird, see Markus Fischer, "A Robot That Flies like a Bird," TED Talk, July 2011, at https://www.ted.com/talks/a_robot_ that_flies_like_a_bird.

الفصل الخامس

1. Ray Kurzweil, *The Singularity Is Near* (New York: Viking Press, 2005).
2. Ben Goertzel's "Nanny AI" scenario is described here: https://wiki.lesswrong .com/wiki/Nanny_AI.
3. For a discussion about the relationship between machines and humans, and whether machines are our slaves, see Benjamin Wallace-Wells, "Boyhood," *New York* magazine (May 20, 2015), online at http://tinyurl.com/aislaves.
4. Mind crime is discussed in Nick Bostrom's book Superintelligence and in more technical detail in this recent paper: Nick Bostrom, Allan Dafoe and Carrick Flynn, "Policy Desiderata in the Development of Machine Superintelligence" (2016), http://www.nickbostrom.com/papers/aipolicy.pdf.
5. Matthew Schofield, "Memories of Stasi Color Germans' View of U.S. Surveillance Programs,"*McClatchy DC Bureau* (June 26, 2013), online at http://www .mcclatchydc.com/news/nation-world/national/article24750439.html.
6. For thought-provoking reflections on how people can be incentivized to create outcomes that nobody wants, I recommend "Meditations on Moloch," http:// slatestarcodex.com/2014/07/30/meditations-on-moloch.

7. For an interactive timeline of close calls when nuclear war might have started by accident, see Future of Life Institute, "Accidental Nuclear War: A Timeline of Close Calls," online at http://tinyurl.com/nukeoops.
8. For compensation payments made to U.S. nuclear testing victims, see U.S. Department of Justice website, "Awards to Date 4/24/2015," at https://www .justice.gov/civil/awards-date-04242015.
9. *Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack*, April 2008, available online at http://www .empcommission.org/docs/ A2473-EMP_Commission-7MB.pdf.
10. Independent research by both U.S. and Soviet scientists alerted Reagan and Gorbachev to the risk of nuclear winter: P. J. Crutzen and J. W. Birks, "TheAtmosphere After a Nuclear War: Twilight at Noon,"Ambio 11, no. 2/3 (1982): 114–125. R. P. Turco, O. B. Toon, T. P. Ackerman, J. B. Pollack and C. Sagan, "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions," Science 222 (1983): 1283–1292. V. V. Aleksandrov and G. L. Stenchikov, "On the Modeling of the Climatic Consequences of the Nuclear War," Proceeding on *Applied Mathematics* (Moscow: Computing Centre of the USSR Academy of Sciences, 1983), 21. A. Robock, "Snow and Ice Feedbacks Prolong Effects of Nuclear Winter," Nature 310 (1984): 667–670.
11. Calculation of climate effects of global nuclear war: A. Robock, L. Oman and L. Stenchikov, "Nuclear Winter Revisited with a Modern Climate Model and Current Nuclear Arsenals: Still Catastrophic Consequences," *Journal of Geophysical Research* 12 (2007): D13107.

الفصل السادس

1. For more information, see Anders Sandberg, "Dyson Sphere FAQ," at http:// www.aleph. se/nada/dysonFAQ.html.
2. Freeman Dyson's seminal paper on his eponymous spheres: Freeman Dyson, "Search for Artificial Stellar Sources of Infrared Radiation," Science, vol. 131 (1959): 1667–1668.
3. Louis Crane and Shawn Westmoreland explain their proposed black hole engine in "Are Black Hole Starships Possible?," at http://arxiv.org/pdf/0908.1803.pdf.
4. For a nice infographic from CERN summarizing known elementary particles, see http:// tinyurl.com/cernparticle.
5. This unique video of a non-nuclear Orion prototype illustrates the idea of nuclear-bomb-powered rocket propulsion: https://www.youtube.com/watch?v =E3Lxx2VAYi8.
6. Here's a pedagogical introduction to laser sailing: Robert L. Forward, "Roundtrip Interstellar Travel Using Laser-Pushed Lightsails," *Journal of Spacecraft and Rockets* 21, no. 2 (March–April 1984), available online at http://www.lunarsail .com/LightSail/rit-1.pdf.
7. Jay Olson analyzes cosmically expanding civilizations in "Homogeneous Cosmology with Aggressively Expanding Civilizations," Classical and Quantum Gravity 32 (2015), available online at http://arxiv.org/abs/1411.4359.
8. The first thorough scientific analysis of our far future: Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (1979): 447, available online at http://blog.regehr.org/ extra_files/dyson.pdf.
9. Seth Lloyd's above-mentioned formula told us that performing a computational operation during a time interval τ costs an energy $E \geq h/4\tau$, where h is Planck's constant. If we want to get N operations done one after the other (in series) during a time T, then $\tau = T/N$, so $E/N \geq hN/4T$, which tells us that we can perform $N \leq 2\sqrt{ET/h}$ serial operations using energy E and time T. So both energy and time are resources that it helps having lots of. If you split your energy between n different parallel computations, they can run more slowly and efficiently, giving $N \leq 2\sqrt{ETn/h}$. Nick Bostrom estimates that simulating a 100-year human life requires about $N = 10^{27}$ operations.

10. If you want to see a careful argument for why the origin of life may require a very rare fluke, placing our nearest neighbors over 10^{1000} meters away, I recommend this video by Princeton physicist and astrobiologist Edwin Turner: "Improbable Life: An Unappealing but Plausible Scenario for Life's Origin on Earth," at https://www.youtube.com/watch?v=Bt6n6Tu1beg.
11. Essay by Martin Rees on the search for extraterrestrial intelligence: https://www .edge. org/annual-question/2016/response/26665.

الفصل السابع

1. A popular discussion of Jeremy England's work on "dissipation-driven adaptation" can be found in Natalie Wolchover, "A New Physics Theory of Life," *Scientific American* (January 28, 2014), available online at https://www.scienti ficamerican.com/article/a-new-physics-theory-of-life/. Ilya Prigogine and Isabelle Stengers's *Order Out of Chaos: Man's New Dialogue with Nature* (New York: Bantam, 1984) lays many of the foundations for this.
2. For more on feelings and their physiological roots: William James, *Principles of Psychology* (New York: Henry Holt & Co., 1890); *Robert Ornstein, Evolution of Consciousness: The Origins of the Way We Think* (New York: Simon & Schuster, 1992); António Damásio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Penguin, 2005); and António Damásio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage, 2012).
3. Eliezer Yudkowsky has discussed aligning the goals of friendly AI not with our present goals, but with our *coherent extrapolated volition* (CEV). Loosely speaking this is defined as what an idealized version of us would want if we knew more, thought faster and were more the people we wished we were. Yudkowsky began criticizing CEV shortly after publishing it in 2004 (http://intelligence.org/files/ CEV.pdf), both for being hard to implement and because it's unclear whether it would converge to anything well-defined.
4. In the inverse reinforcement-learning approach, a core idea is that the AI is trying to maximize not its own goal-satisfaction, but that of its human owner. It therefore has incentive to be cautious when it's unclear about what its owner wants, and to do its best to find out. It should also be fine with its owner switching it off, since that would imply that it had misunderstood what its owner really wanted.
5. Steve Omohundro's paper on AI goal emergence, "The Basic AI Drives," can be found online at http://tinyurl.com/omohundro2008. Originally published in *Artificial General Intelligence 2008: Proceedings of the First AGI Conference,* ed. Pei Wang, Ben Goertzel and Stan Franklin (Amsterdam: IOS, 2008), 483–492.
6. A thought-provoking and controversial book on what happens when intelligence is used to blindly obey orders without questioning their ethical basis: Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin, 1963). A related dilemma applies to a recent proposal by Eric Drexler (http://www.fhi.ox.ac.uk/reports/2015-3.pdf) to keep superintelligence under control by compartmentalizing it into simple pieces, none of which understand the whole picture. If this works, this could again provide an incredibly powerful tool without an intrinsic moral compass, implementing its owner's every whim without any moral qualms. This would be reminiscent of a compartmentalized bureaucracy in a dystopian dictatorship: one part builds weapons without knowing how they'll be used, another executes prisoners without knowing why they were convicted, and so on.
7. A modern variant of the Golden Rule is John Rawls' idea that a hypothetical situation is fair if nobody would change it without knowing in advance which person in it they'd be.
8. For example, the IQs of many of Hitler's top officials were found to be quite high. See "How Accurate Were the IQ Scores of the High-Ranking Third Reich Officials Tried at Nuremberg?," Quora, available online at http://tinyurl .com/nurembergiq.

الفصل الثامن

1. The entry on consciousness by Stuart Sutherland is quite amusing: *Macmillan Dictionary of Psychology* (London: Macmillan, 1989).
2. Erwin Schrödinger, one of the founding fathers of quantum mechanics, made this remark in his book *Mind and Matter* while contemplating the past—and what would have happened if conscious life never evolved in the first place. On the other hand, the rise of AI raises the logical possibility that we may end up with a play for empty benches in the future.
3. The *Stanford Encyclopedia of Philosophy* gives an extensive survey of different definitions and uses of the word "consciousness": http://tinyurl.com/stanford consciousness.
4. Yuval Noah Harari, *Homo Deus: A Brief History of Tomorrow* (New York: HarperCollins, 2017): 116.
5. This is an excellent introduction to System 1 and System 2 from a pioneer: Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus & Giroux, 2011).
6. See Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (New York: W. H. Freeman, 2004).
7. Perhaps we're only aware of a tiny fraction (say 10–50 bits) of the information that enters our brain each second: K. Küpfmüller, 1962, "Nachrichtenverarbeitung im Menschen," in *Taschenbuch der Nachrichtenverarbeitung*, ed. K.Steinbuch (Berlin: Springer-Verlag, 1962): 1481–1502. T. Nørretranders, *The User Illusion: Cutting Consciousness Down to Size* (New York: Viking, 1991).
8. Michio Kaku, *The Future of the Mind: The Scientific Quest to Understand, Enhance, and Empower the Mind* (New York: Doubleday, 2014); Jeff Hawkins and Sandra Blakeslee, On Intelligence (New York: Times Books, 2007); Stanislas Dehaene, Michel Kerszberg and Jean-Pierre Changeux, "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks," *Proceedings of the National Academy of Sciences* 95 (1998): 14529–14534.
9. Video celebrating Penfield's famous "I can smell burnt toast" experiment: https: //www.youtube.com/watch?v=mSN86kphL68. Sensorimotor cortex details: Elaine Marieb and Katja Hoehn, *Anatomy & Physiology*, 3rd ed. (Upper Saddle River, NJ: Pearson, 2008), 391–395.
10. The study of neural correlates of consciousness (NCCs) has become quite mainstream in the neuroscience community in recent years—see, e.g., Geraint Rees, Gabriel Kreiman, and Christof Koch, "Neural Correlates of Consciousness in Humans," Nature Reviews Neuroscience 3 (2002): 261–270, and Thomas Metzinger, *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Cambridge, MA: MIT Press, 2000).
11. How continuous flash suppression works: Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (New York: W. H. Freeman, 2004); Christof Koch and Naotsugu Tsuchiya, "Continuous Flash Suppression Reduces Negative Afterimages," *Nature Neuroscience* 8 (2005): 1096–1101.
12. Christof Koch, Marcello Massimini, Melanie Boly and Giulio Tononi, "Neural Correlates of Consciousness: Progress and Problems," Nature Reviews Neuroscience 17 (2016): 307.
13. See Koch, *The Quest forConsciousness*, p. 260, and further discussion in *the Stanford Encyclopedia of Philosophy*, http://tinyurl.com/consciousnessdelay.
14. On synchronization of conscious perception: David Eagleman, *The Brain: The Story of You* (New York: Pantheon, 2015), and *Stanford Encyclopedia of Philosophy*, http://tinyurl.com/consciousnesssync.
15. Benjamin Libet, *Mind Time: The Temporal Factor in Consciousness* (Cambridge, MA: Harvard University Press, 2004); Chun Siong Soon, Marcel Brass, HansJochen Heinze and John-Dylan Haynes, "Unconscious Determinants of Free Decisions in the Human Brain," Nature Neuroscience 11 (2008): 543–545, online at http://www.nature.com/neuro/journal/v11/n5/full/nn.2112.html.

16. Examples of recent theoretical approaches to consciousness:

    Daniel Dennett, *Consciousness Explained* (Back Bay Books, 1992)

    Bernard Baars, *In the Theater of Consciousness: The Workspace of the Mind* (New York: Oxford University Press, 2001)

    Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (New York: W. H. Freeman, 2004)

    Gerald Edelman and Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Hachette, 2008)

    António Damásio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage, 2012)

    Stanislas Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (New York: Viking, 2014)

    Stanislas Dehaene, Michel Kerszberg and Jean-Pierre Changeux, "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks," *Proceedings of the National Academy of Sciences* 95 (1998): 14529–14534

    Stanislas Dehaene, Lucie Charles, Jean-Rémi King and Sébastien Marti, "Toward a Computational Theory of Conscious Processing," *Current Opinion in Neurobiology* 25 (2014): 760–784

17. Thorough discussion of different uses of the term "emergence" in physics and philosophy by David Chalmers: http://cse3521.artifice.cc/Chalmers -Emergence.pdf.
18. Me arguing that consciousness is the way information feels when being processed in certain complex ways: https://arxiv.org/abs/physics/0510188, https:// arxiv.org/abs/0704.0646, Max Tegmark, Our Mathematical Universe (New York: Knopf, 2014). David Chalmers expresses a related sentiment in his 1996 book The Conscious Mind: "Experience is information from the inside; physics is information from the outside."
19. Adenauer Casali et al., "A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior," Science Translational Medicine 5 (2013): 198ra105, online at http://tinyurl.com/zapzip.
20. Integrated information theory doesn't work for continuous systems: - https://arxiv.org/abs/1401.1219 - http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00063/full - https://arxiv.org/abs/1601.02626
21. Interview with Clive Wearing, whose short-term memory is only about 30 seconds: https://www.youtube.com/watch?v=WmzU47i2xgw.
22. Scott Aaronson IIT critique: http://www.scottaaronson.com/blog/?p=1799.
23. Cerrullo IIT critique, arguing that integration isn't a sufficient condition for consciousness: http://tinyurl.com/cerrullocritique.
24. IIT prediction that simulated humans will be zombies: http://rstb.royalsociety publishing.org/content/370/1668/20140167.
25. Shanahan critique of IIT: https://arxiv.org/pdf/1504.05696.pdf.
26. Blindsight: http://tinyurl.com/blindsight-paper.
27. Perhaps we're only aware of a tiny fraction (say 10–50 bits) of the information that enters our brain each second: Küpfmüller, "Nachrichtenverarbeitung im Menschen"; Nørretranders, *The User Illusion.*
28. The case for and against "consciousness without access": Victor Lamme, "How Neuroscience Will Change Our View on Consciousness," *Cognitive Neuroscience* (2010): 204–220, online at http://www.tandfonline.com/doi/abs/10 .1080/17588921003731586.
29. "Selective Attention Test," at https://www.youtube.com/watch?v=vJG698U2 Mvo.
30. See Lamme, "How Neuroscience Will Change Our View on Consciousness," n. 28.
31. This and other related issues are discussed in detail in Daniel Dennett's book *Consciousness Explained.*

32. See Kahneman, Thinking, Fast and Slow, cited in n. 5.
33. The Stanford Encyclopedia of Philosophy reviews the free will controversy: https:// plato. stanford.edu/entries/freewill.
34. Video of Seth Lloyd explaining why an AI will feel like it has free will: https:// www.youtube. com/watch?v=Epj3DF8jDWk.
35. See Steven Weinberg, *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature* (New York: Pantheon, 1992).
36. The first thorough scientific analysis of our far future: Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (1979): 447, available online at http://blog.regehr.org/ extra_files/dyson.pdf.

الخاتمة

1. The open letter (http://futureoflife.org/ai-open-letter) that emerged from the Puerto Rico conference argued that research on how to make AI systems robust and beneficial is both important and timely, and that there are concrete research directions that can be pursued today, as exemplified in this research-priorities document: http://futureoflife.org/ data/documents/research_priorities.pdf.
2. My video interview with Elon Musk about AI safety can be found on YouTube at https:// www.youtube.com/watch?v=rBw0eoZTY-g.
3. Here's a nice video compilation of almost all SpaceX rocket landing attempts, culminating with the first successful ocean landing: https://www.youtube.com/ watch?v=AllaFzIPaG4.
4. Elon Musk tweets about our AI-safety grant competition: https://twitter.com/ elonmusk/ status/555743387056226304.
5. Elon Musk tweets about our AI-safety-endorsing open letter: https://twitter .com/elonmusk/ status/554320532133650432.
6. Erik Sofge in "An Open Letter to Everyone Tricked into Fearing Artificial Intelligence" (*Popular Science*, January 14, 2015) pokes fun at the scaremongering news coverage of our open letter: http://www.popsci.com/open-letter -everyone-tricked-fearing-ai.
7. Elon Musk tweets about his big donation to the Future of Life Institute and the world of AI-safety researchers: https://twitter.com/elonmusk/status/55574338 7056226304.
8. For more about the Partnership on AI to benefit people and society, see their website: https://www.partnershiponai.org.
9. Some examples of recent reports expressing opinions about AI: One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Panel, "Artificial Intelligence and Life in 2030" (September 2016), at http://tinyurl.com/ stanfordai; White House report on the future of AI: http://tinyurl.com/obama AIreport; White House report on AI and jobs: http://tinyurl. com/AIjobsreport; IEEE report on AI and human well-being, "Ethically Aligned Design, Version1" (December 13, 2016), at http://standards.ieee.org/develop/indconn/ec/ead_v1 .pdf; road map for U.S. Robotics: http://tinyurl.com/roboticsmap.
10. Among the principles that didn't make the final cut, one of my favorites was this one: "Consciousness caution: There being no consensus, we should avoid strong assumptions as to whether or not advanced AI would possess or require consciousness or feelings." It went through many iterations, and in the last one, the controversial word "consciousness" was replaced by "subjective experience"— but this principle nonetheless got only 88% approval, just barely falling short of the 90% cutoff.
11. Discussion panel on superintelligence with Elon Musk and other great minds: http://tinyurl. com/asilomarAI.

حول المؤلف

ماكس تيغمارك Tegmark هو أستاذ الفيزياء في المعهد MIT ورئيس معهد مستقبل الحياة Future of Life Institute. وهو مؤلف كتاب **كوننا الرياضياتي** ***Our Mathematical Universe***. وقد ظهر في عشرات من الأفلام الوثائقية العلمية. وهو شغوف بالأفكار والمغامرة ومستقبل مُلهِم مُعدٍ.